Tutorial:
Using the JAR3D website to score sequences of RNA internal and hairpin loops against RNA 3D motif groups

Jim Roll, Craig Zirbel, and the rest of the Bowling Green State University RNA group

Last updated December 31, 2015

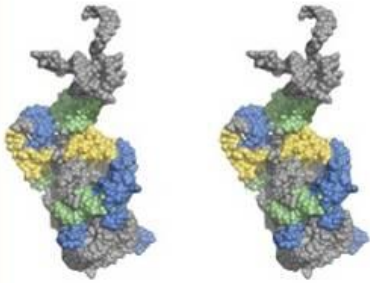# Table of contents - hyperlinked

Overview of RNA 3D structure

RNA is hierarchical, with primary structure (the sequence of A, C, G, and U nucleotides), secondary structure (the pattern of Watson-Crick basepairs AU, GC, and GU that form RNA helices joined by junction and internal "loops" and capped by hairpin "loops"), tertiary structure (how the helical and loop regions assemble in three dimensions), and quaternary structure (interactions between an RNA molecule and other RNA or proteins).
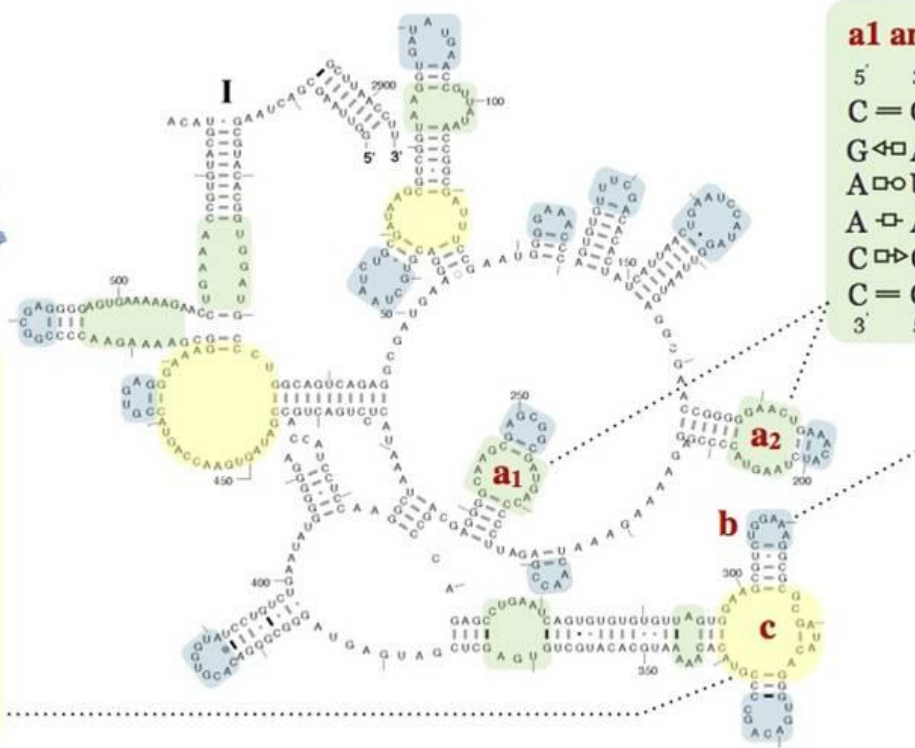
RNA secondary structures are inferred by two main methods, **folding** and **covariation**. Folding starts with a single sequence of RNA nucleotides and tries to pair up subsequences that make the most AU and GC basepairs. Standard programs that fold RNA are mfold (now being replaced by UNAFold), RNAfold, and also LocaRNA when multiple sequences are available. Covariation uses sequences of the same RNA molecule from related organisms and looks for changes between AU, GC, CG, and UA in the positions that are likely to be making Watson-Crick basepairs. Usually, folding and covariation analysis have to be done together. A standard source for predicted RNA secondary structures is the Gutell Lab and in particular the Current Structure Models for Reference Organisms.
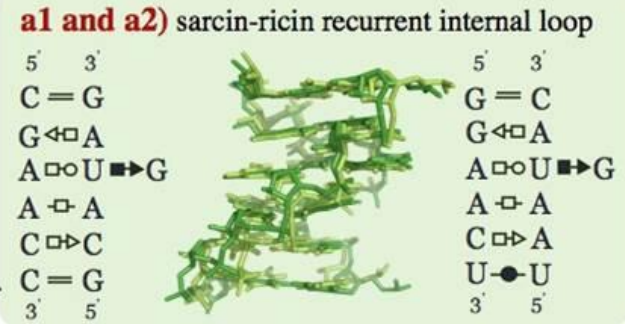
RNA primary, secondary, and tertiary structure

RNA primary, secondary, and tertiary structure. The center image is the secondary structure of Domain 1 of the large ribosomal subunit. The primary sequence starts in the upper left where it reads 5'. Nucleotides are numbered sequentially from there, with a line marking every 10th nucleotide and numbers marking every 50th. RNA helices appear as rectangular blocks of ACGU characters connected by straight lines, which represent the Watson-Crick basepairs.

RNA hairpins are shaded blue, they cap helical regions. RNA internal loops are shaded green, they connect two helices. RNA junctions connect three or more helices, they are shaded yellow, except for the central junction which connects seven helices and is shown as a large circle in the middle of the diagram, and the smaller junction down and to the left of there, which connects four helices.

RNA primary, secondary, and tertiary structure

Stereo view of 3D structure

2D diagram

RNA 3D motifs with 2D annotations

a1 and a2) sarcin-ricin recurrent internal loop

c) three-way junction

b) T-loop hairpin
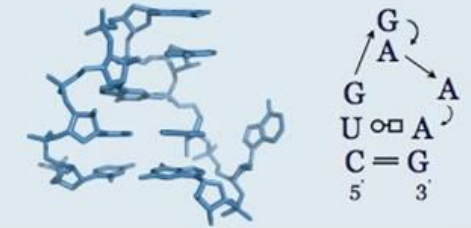
RNA hairpins, internal loops, and junctions. The inset figures show the 3D structures of selected hairpins, internal loops, and junctions. The internal and hairpin loops are annotated with basepair diagrams that show the non-Watson-Crick basepairs using additional symbols.
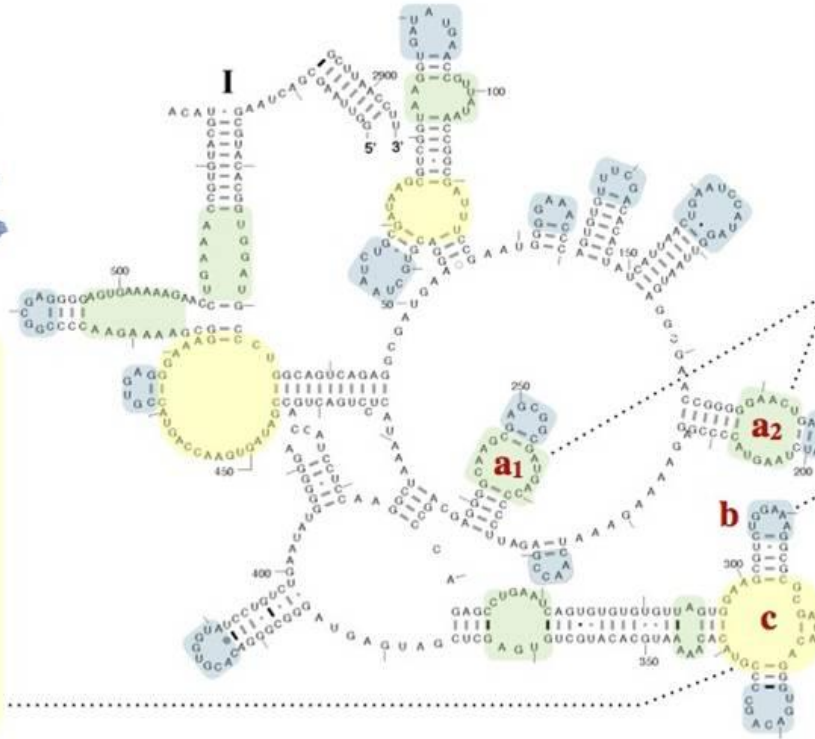
The two internal loops shown have essentially the same 3D structure and have the same basepair diagram but have slightly different sequences. We find that the geometry and the pattern of basepairs that occur in loop regions are often recurrent, appearing in essentially the same form in multiple RNA molecules not related by evolution.

The key to the tool being introduced in this tutorial is that non-Watson-Crick basepairs have their own characteristic patterns of sequence variability. Having observed the non-Watson-Crick basepairs that structure a "loop" region, we can tell what sequence variations or covariations will be consistent with that loop, and so identify additional instances of that loop that have a different sequence from what was observed in a 3D structure.

How to extract the sequence of an internal loop

Extracting sequences of loops for JAR3D input: Refer to the secondary structure below. Find the internal loop starting at position 580 (blue box). It connects two RNA helices, which appear as nucleotides connected by straight lines. Identify the flanking AU, GC, or GU basepairs. These are the last Watson-Crick basepairs of the two helices on either side of the internal loop. Write down the sequence of nucleotides, starting with one flanking Watson-Crick pair and continuing in order of increasing nucleotide number (i.e. in 5' to 3' order), up to the second flanking Watson-Crick pair. Then write down an asterisk, *, to separate the letters of the two strands making up the internal loop. Continue by writing down the nucleotides of the second strand, from one flanking Watson-Crick pair to the other. For this example, the result should be "CGCAG*UCAGG" or, starting from the other Watson-Crick basepair, "UCAGG*CGCAG."
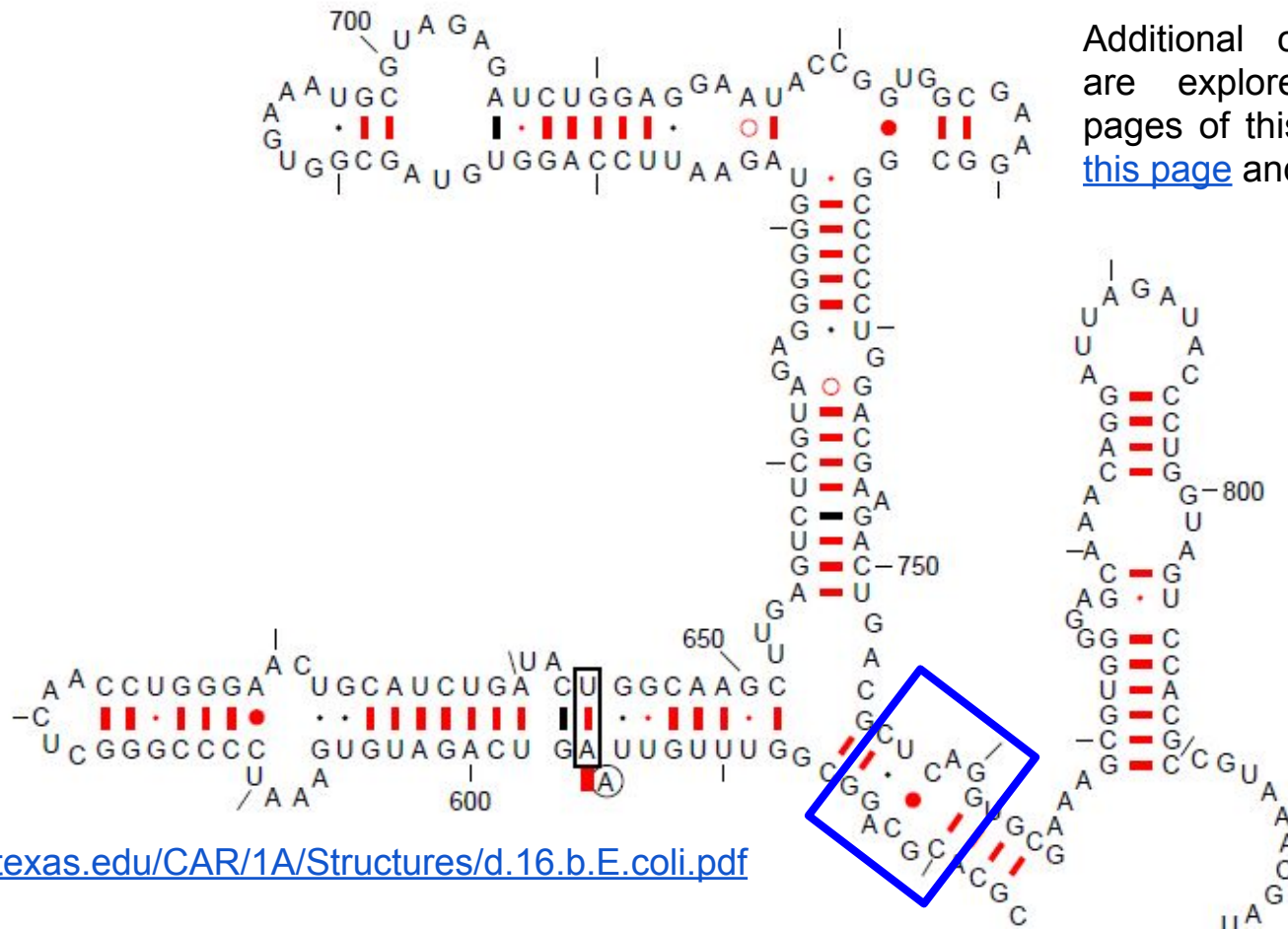
Additional considerations are explored on later pages of this tutorial, see this page and this page.

www.rna.ccbb.utexas.edu/CAR/1A/Structures/d.16.b.E.coli.pdf

Overview and history of JAR3D

JAR3D is a tool to address the following problem: Given one or more sequences of an RNA internal or hairpin loop, what known 3D geometries are consistent with the sequence(s)? Answering this question is a step on the way to determining the 3D structure of an RNA molecule. The JAR3D webserver is the culmination of a long line of research which we briefly review here.

First, there are characteristic ways in which RNA internal and hairpin loops are structured. The Leontis-Westhof system for annotating 12 families of Watson-Crick and non-Watson-Crick basepairs is the key step here. (Leontis, Stombaugh, Westhof 2002) Base-phosphate interactions also play important roles in structuring loops. (Zirbel, Sponer, Sponer, Stombaugh, Leontis 2009) Stacking interactions are also important, but we have yet to understand the base specificity of stacking interactions in general.

Next, each basepair in a structured RNA can change base combination due to DNA copying errors while maintaining the same basic geometry. This is familiar in RNA helices where AU changes to CG, but similar changes also occur in non-Watson-Crick basepairs in RNA loops. The propensity to change is measured using basepair isostericity, and detailed studies of homologous RNA molecules supports the idea that isosteric changes occur most often. (Stombaugh, Zirbel, Westhof, Leontis 2009)

Many RNA loops are recurrent in the sense that they appear with the same geometry and the same pattern of basepairing in non-homologous places in different RNAs, and often with different sequences. This is the reason that it makes sense to ask whether a new RNA loop sequence can match a known 3D geometry. The RNA 3D Motif Atlas was developed to group together instances of RNA internal and hairpin loops that share the same basic geometry and pattern of basepairing. The Atlas is updated as additional RNA structures are solved, thus improving the chance of finding a match to a new sequence. (Petrov, Zirbel, Leontis 2013) To avoid duplication from the many solved structures of the same molecule, for example E. coli small ribosomal subunit, we developed non-redundant sets of RNA 3D structures. (Leontis and Zirbel 2012)

Finally, we need a way to score the sequence of an RNA internal or hairpin loop against a motif group to tell whether it is a plausible match or not. JAR3D was developed to model sequence variability in RNA motifs and to align sequences to motif groups. (Zirbel, Roll, Sweeney, Petrov, Pirrung, Leontis 2015) JAR3D stands for "Java-based Alignment of RNA using 3D structure." Sequences can be scored against each motif group from the RNA 3D Motif Atlas (over 200 IL groups and over 200 HL groups). An acceptance region is defined for each motif group to reduce false positive matches, and a cutoff score gauges how well a sequence matches a motif group, ranging from a minimum of 0 at the edge of the acceptance region to 100 for perfect matches.

JAR3D input page

Submitting input to JAR3D: On the previous slide we extracted this sequence of an internal loop from 16S rRNA: CGCAG*UCAGG. The asterisk (" * ") character denotes the strand break in the internal loop. This slide shows how to submit this sequence to JAR3D (URL: http://rna.bgsu.edu/jar3d/). The screen shot shows an example of the JAR3D input page, with a search for this IL sequence entitled "JAR3D Tutorial Search" being run on version 1.18 of the Motif Atlas. Click here for the direct link to the input page. The links to the right of the input box show how to input multiple sequences of one loop or sequences with multiple loops and a secondary structure. Examples of each of these will be discussed later.

JAR3D Scoring RNA loop sequences against known 3D motifs

Tutorial

JAR3D Tutorial Search

CGCAG*UCAGG

Submit   Clear   Motif Atlas Version : 1.18

Sample Input
Single sequence, one loop
Multiple sequences, one loop
Multiple sequences, one loop (fasta)
Single sequence, many loops, with secondary structure
Single sequence, many loops, with secondary structure (fasta)
Multiple sequences, many loops, with secondary structure
Multiple sequences, many loops, with secondary structure (fasta)

Goal

JAR3D scores RNA hairpin and internal loop sequences against motif groups from the RNA 3D Motif Atlas, by exact sequence match for sequences already observed in 3D and by probabilistic scoring and edit distance for novel sequences. RNA hairpin and internal loops are often represented on secondary structure diagrams as if they are unstructured, but in fact most are structured by non-Watson-Crick basepairs, base stacking, and base-backbone interactions. Analysis of 3D structures shows that different RNA sequences can form the same RNA 3D motif, as is apparent in many motif groups in the

Input and Output

JAR3D accepts single or multiple sequences having one or many loops. See the Examples above. **One loop:** To specify the break between strands in internal loops, use an asterisk *. Sequence(s) without an asterisk are interpreted as hairpins. Internal and hairpin loops should include closing Watson-Crick basepairs, with nucleotides running in 5' to 3' order within each strand. Individual loops do not need the nucleotides to be aligned. **Many loops:** JAR3D will extract internal and hairpin loops from longer sequences if a dot-bracket secondary structure is provided as the first line of the input. Multiple

Method

1. We extract all hairpin and internal loops from a non-redundant set of RNA 3D structures from the PDB/NDB and cluster them in geometrically similar families.

2. For each recurrent motif, we construct a probabilistic model for sequence variability based on a hybrid **Stochastic Context-Free Grammar/Markov Random Field** (SCFG/MRF) method we developed.

3. To parameterize each model, we use all instances of the motif found in the non-redundant dataset and knowledge of RNA nucleotide interactions, especially isosteric basepairs

JAR3D output: Input summary, column numbers

The JAR3D output page begins with the name given to the JAR3D search ("JAR3D Tutorial Search") followed by a summary of the input. Because multiple loops can be input, loops are numbered and listed one by one. Under the heading "Loop 1" is a list of the column numbers in the original input where Loop 1 appears. This is not needed for a single loop as shown here, but is very helpful when a longer RNA sequence with multiple loops is input.

JAR3D Query 4295f189-9518-47f9-a4bb-e3d013c3c4f9 completed
Run on Motif Atlas Version 1.18

➕Share

JAR3D Tutorial Search
Input Summary

Column numbers

CGCAG*UCAGG

Refine query

Loop 1
Column Numbers: 1, 2, 3, 4, 5, 6, 7, 8, 9

CGCAG*UCAGG

Click on headings to reorder table

#	Motif ID and Additional Information	Acceptance Rate	Mean Cutoff Score	Full Edit Distance	Interior Edit Distance	2D Diagram
1	⦿ IL_77895.1 View IL_77895.1 cWW-tSH-tWH-tHS-cWW Align sequences to IL_77895.1	100.00	80.45	0	0	
2	○ IL_98284.1 View IL_98284.1 cWW-cWW-tSH-tWH-tHS-cWW Align sequences to IL_98284.1	100.00	15.61	2	2	
3	○ IL_64847.2 View IL_64847.2 cWW-tSH-tHW-tHS-cWW Align sequences to IL_64847.2	100.00	14.27	2	2	

Show neighborhood   Show stereo   Previous   Next
☐ Numbers

[Persistent link to this result page](#)

JAR3D output: Matching motif groups

JAR3D output for input sequence CGCAG*UCAGG is shown on this slide. Each row of the table tells how this sequence matches a different possible internal loop motif group from the RNA 3D Motif Atlas. Motif groups are identified by "IL" or "HL" followed by a 5-digit number followed by a version number after the period. Links to the web page for each motif group in the RNA 3D Motif Atlas are provided. Up to ten motif groups will be listed.

[Persistent link to this result page](#)

JAR3D output: Motif basepair signatures

The basepair signature gives a quick indication of consensus basepairs present in the given motif group, using the Leontis-Westhof notation, in which "t" stands for "trans," "c" for "cis," "W," for Watson-Crick edge, "H" for "Hoogsteen edge," and "S" for "Sugar edge." See the RNA basepair catalog for examples of each basepair.

JAR3D output: Motif basepair diagrams

Basepair diagrams also indicate the consensus basepairs present in the given motif group, using the Leontis-Westhof symbols in which a circle stands for Watson-Crick edge, square for Hoogsteen edge, triangle for sugar edge, open symbol for trans, and closed symbol for cis. The basepair diagram for the first motif group below suggests that it consists of five stacked basepairs, and the 3D view confirmsthat this is the case.

Basepair diagrams

[Persistent link to this result page](#)

JAR3D output: 3D structure view

Radio buttons can be used to select a motif group. The jsmol window on the right shows the 3D structure of one instance from the selected motif group. It can be rotated or zoomed by clicking and dragging. Notice that the selected motif comprises three stacked non-Watson-Crick basepairs, as suggested by the basepair diagram. Use the persistent link to view and rotate in 3D.

JAR3D Query 4295f189-9518-47f9-a4bb-e3d013c3c4f9 completed
Run on Motif Atlas Version 1.18

➕Share

JAR3D Tutorial Search
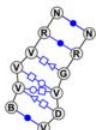### Input Summary

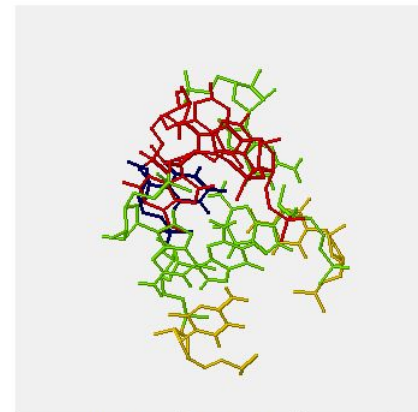CGCAG*UCAGG

Refine query

Loop 1
Column Numbers: 1, 2, 3, 4, 5, 6, 7, 8, 9

CGCAG*UCAGG

Click on headings to reorder table

#	Motif ID and Additional Information	Acceptance Rate	Mean Cutoff Score	Full Edit Distance	Interior Edit Distance	2D Diagram
1	◉ IL_77895.1 View IL_77895.1 cWW-tSH-tWH-tHS-cWW Align sequences to IL_77895.1	100.00	80.45	0	0	
2	○ IL_98284.1 View IL_98284.1 cWW-cWW-tSH-tWH-tHS-cWW Align sequences to IL_98284.1	100.00	15.61	2	2	
3	○ IL_64847.2 View IL_64847.2 cWW-tSH-tHW-tHS-cWW Align sequences to IL_64847.2	100.00	14.27	2	2	

Show neighborhood Show stereo Previous Next
☐ Numbers

Persistent link to this result page

JAR3D output: Diagnostic Information (DI)

Each row of the output table lists the percentage of input sequences falling in the acceptance region for the motif group, then the mean cutoff score. In addition, the median minimum edit distance to a sequence from 3D is shown, both for the full sequences and the interior sequence (excluding flanking basepairs). An edit distance of 0 indicates an exact sequence match. Motif groups are listed by acceptance rate and then by cutoff score.

JAR3D Query 4295f189-9518-47f9-a4bb-e3d013c3c4f9 completed

Run on Motif Atlas Version 1.18

+Share

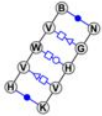## JAR3D Tutorial Search
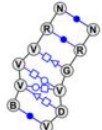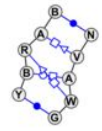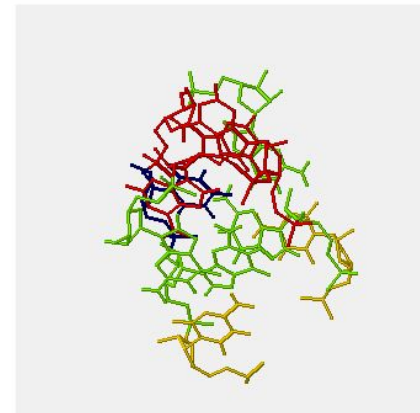**Input Summary**

CGCAG*UCAGG

Refine query

Loop 1
Column Numbers: 1, 2, 3, 4, 5, 6, 7, 8, 9

CGCAG*UCAGG

Click on headings to reorder table

#	Motif ID and Additional Information	Acceptance Rate	Mean Cutoff Score	Full Edit Distance	Interior Edit Distance	2D Diagram
1	⦿ IL_77895.1 View IL_77895.1 cWW-tSH-tWH-tHS-cWW Align sequences to IL_77895.1	100.00	80.45	0	0	
2	⦾ IL_98284.1 View IL_98284.1 cWW-cWW-tSH-tWH-tHS-cWW Align sequences to IL_98284.1	100.00	15.61	2	2	
3	⦾ IL_64847.2 View IL_64847.2 cWW-tSH-tHW-tHS-cWW Align sequences to IL_64847.2	100.00	14.27	2	2	

Show neighborhood   Show stereo   Previous   Next
☐ Numbers

Cutoff scores above 0 fall in the acceptance region. Cutoff scores of 100 indicate the best match of a sequence to the motif group. Negative cutoff scores indicate a poor match.

More information on acceptance regions and cutoff scores can be found in the JAR3D article here.

Persistent link to this result page

JAR3D output: Evaluating matches

The first row of the results table shows a good match for the query sequence. The sequence falls in the acceptance region, it has a high cutoff score (near 100), and it has full and interior edit distance of zero, meaning it is an exact match to a sequence that has been seen in 3D. To find out which 3D instance it matches exactly, click on Align Sequences.

JAR3D Query 4295f189-9518-47f9-a4bb-e3d013c3c4f9 completed

Run on Motif Atlas Version 1.18

➕Share

JAR3D Tutorial Search

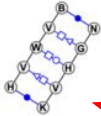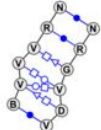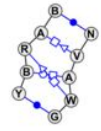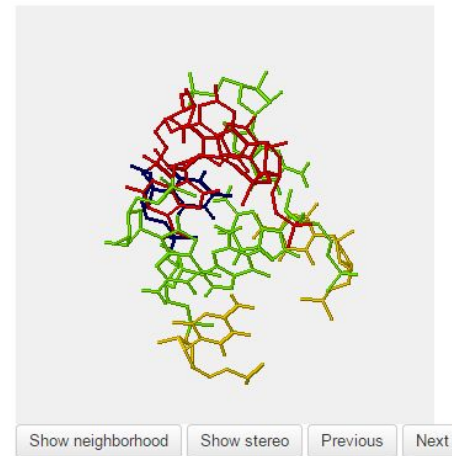### Input Summary

CGCAG*UCAGG

Refine query

Loop 1

Column Numbers: 1, 2, 3, 4, 5, 6, 7, 8, 9

CGCAG*UCAGG

Click on headings to reorder table

#	Motif ID and Additional Information	Acceptance Rate	Mean Cutoff Score	Full Edit Distance	Interior Edit Distance	2D Diagram
1	⦿ IL_77895.1 View IL_77895.1 cWW-tSH-tWH-tHS-cWW Align sequences to IL_77895.1	100.00	80.45	0	0	
2	⦾ IL_98284.1 View IL_98284.1 cWW-cWW-tSH-tWH-tHS-cWW Align sequences to IL_98284.1	100.00	15.61	2	2	
3	⦾ IL_64847.2 View IL_64847.2 cWW-tSH-tHW-tHS-cWW Align sequences to IL_64847.2	100.00	14.27	2	2	

Show neighborhood   Show stereo   Previous   Next
☐ Numbers

[Persistent link to this result page](#)

JAR3D alignment to a motif group

This JAR3D alignment page shows an alignment of the input sequence CGCAG*UCAGG (gray background) to sequences of 3D instances from motif group IL_77895.1 (white background). The 3D instances are aligned to one another in 3D as part of the creation of the RNA 3D Motif Atlas. The Position number in the header indicates the conserved positions in the 3D motif group. Note that 3D instance IL_3J7A_061 has an inserted G between Position 3 and Position 4. DI for each input sequence is provided in the four columns to the right of the alignment.

JAR3D Query 4295f189-9518-47f9-a4bb-e3d013c3c4f9   **+Share**

Alignment of Loop 0 to IL_77895.1

Run on Motif Atlas Version 1.18

Input sequences are initially listed on the top, and sequences from 3D structures on the bottom. Click on a column header to sort by that column. Holding shift and clicking on multiple column headers will sort by each column in order.
The strands of your internal loop have been switched to match the strand ordering seen in IL_77895.1. They are still in 5' to 3' order.

Column Position Insertion	1	2 1	3	4 2	5	6 3	7	8 4	9	10 5	11	12	13	14 6	15	16 7	17	18 8	19	20 9	21	22 10	23	Meets Cutoff	Cutoff Score	Interior Edit Distance	Full Edit Distance	Alignment Distance to Sequence0
		I		I		I		I		I		I		I		I		I		I		I						
Sequence0		U		C		A		G		G		*		C		G		C		A		G		True	80.446070	0	0	0
IL_1FJG_023		U		G		A		G		G				U		G		U		A		G						3
IL_1FJG_033		C		A		A		A		C				G		G		U		A		G						6
IL_1S72_023		C		G		A		G		G				C		G		C		A		G						2
IL_1S72_055		U		C		A		A		C				G		G		U		C		G						5
IL_2AW7_030		U		C		A		G		G				C		G		C		A		G						0
IL_2AW7_040		C		A		A		A		C				G		G		U		A		G						6
IL_2QBG_052		C		G		A		G		G				C		G		U		A		G						3
IL_2QBG_058		U		G		A		G		G				U		G		U		A		G						3
IL_2QWY_004		A		A		A		A		G				C		G		U		A		U						5
IL_3J7A_029		U		C		A		A		G				C		G		U		A		G						2
IL_3J7A_061		C		C		U	G	C		U				A		G		A		G		G						8
IL_3U5F_035		U		C		A		A		G				C		G		U		A		G						2

[Persistent link to this alignment page](#)

JAR3D alignment: Strand order

The original sequence we input to JAR3D was CGCAG*UCAGG, but it matches positions 1 to 10 in motif group IL_77895.1 with the strands reversed, meaning UCAGG*CGCAG. Thus, the strand order of the input sequence has been switched for this alignment, but the sequence within each strand is unaffected. RNA internal loops are both recurrent and modular; they can join two helices using two different strand orders in different RNA molecules. See also this page of the tutorial.

JAR3D Query 4295f189-9518-47f9-a4bb-e3d013c3c4f9   +Share

Alignment of Loop 0 to IL_77895.1

Run on Motif Atlas Version 1.18

Input sequences are initially listed on the top, and sequences from 3D structures on the bottom. Click on a column header to sort by that column. Holding shift and clicking on multiple column headers will sort by each column in order.

The strands of your internal loop have been switched to match the strand ordering seen in IL_77895.1. They are still in 5' to 3' order.

Column Position Insertion	1	2 1 I	3 2	4 3 I	5 I	6 3 I	7	8 4 I	9	10 5 I	11	12 I	13	14 6 I	15	16 7 I	17	18 8 I	19	20 9 I	21	22 10	23 I	Meets Cutoff	Cutoff Score	Interior Edit Distance	Full Edit Distance	Alignment Distance to Sequence0
Sequence0	I	U	C	A		G		G		*		C		G		C		A		G				True	80.446070	0	0	0
IL_1FJG_023		U	G	A		G		G				U		G		U		A		G								3
IL_1FJG_033		C	A	A		A		C				G		G		U		A		G								6
IL_1S72_023		C	G	A		G		G				C		G		C		A		G								2
IL_1S72_055		U	C	A		A		C				G		G		U		C		G								5
IL_2AW7_030		U	C	A		G		G				C		G		C		A		G								0
IL_2AW7_040		C	A	A		A		C				G		G		U		A		G								6
IL_2QBG_052		C	G	A		G		G				C		G		U		A		G								3
IL_2QBG_058		U	G	A		G		G				U		G		U		A		G								3
IL_2QWY_004		A	A	A		A		G				C		G		U		A		U								5
IL_3J7A_029		U	C	A		A		G				C		G		U		A		G								2
IL_3J7A_061		C	C	U	G	C		U				A		G		A		G		G								8
IL_3U5F_035		U	C	A		A		G				C		G		U		A		G								2

Persistent link to this alignment page

JAR3D alignment: Edit distance to known instances

Cells in the column(s) labeled "Alignment Distance" are colored according to the edit distance from the sequence named in the column header to the sequence in each row. Alignment Distance 0 in the last column shows that the input sequence is an exact match to the 5th sequence from 3D. Larger edit distances are colored with darker shades of red. This column is sortable by clicking on it, as shown on the next slide.

JAR3D Query 4295f189-9518-47f9-a4bb-e3d013c3c4f9   ➕Share

Alignment of Loop 0 to IL_77895.1

Run on Motif Atlas Version 1.18

Input sequences are initially listed on the top, and sequences from 3D structures on the bottom. Click on a column header to sort by that column. Holding shift and clicking on multiple column headers will sort by each column in order.
The strands of your internal loop have been switched to match the strand ordering seen in IL_77895.1. They are still in 5' to 3' order.

Column Position Insertion	1		2 1 I	3	4 2 I	5	6 3 I	7	8 4 I	9	10 5 I	11	12	13 I	14 6	15 I	16 7	17 I	18 8	19 I	20 9	21 I	22 10	23 I	Meets Cutoff	Cutoff Score	Interior Edit Distance	Full Edit Distance	Alignment Distance to Sequence0
Sequence0			U	C	A		G		G			*		C		G		C		A		G				True	80.446070	0	0	0
IL_1FJG_023			U	G	A		G		G					U		G		U		A		G								3
IL_1FJG_033			C	A	A		A		C					G		G		U		A		G								6
IL_1S72_023			C	G	A		G		G					C		G		C		A		G								2
IL_1S72_055			U	C	A		A		C					G		G		U		C		G								5
IL_2AW7_030			U	C	A		G		G					C		G		C		A		G								0
IL_2AW7_040			C	A	A		A		C					G		G		U		A		G								6
IL_2QBG_052			C	G	A		G		G					C		G		U		A		G								3
IL_2QBG_058			U	G	A		G		G					U		G		U		A		G								3
IL_2QWY_004			A	A	A		A		G					C		G		U		A		U								5
IL_3J7A_029			U	C	A		A		G					C		G		U		A		G								2
IL_3J7A_061			C	C	U	G	C		U					A		G		A		G		G								8
IL_3U5F_035			U	C	A		A		G					C		G		U		A		G								2

[Persistent link to this alignment page](#)

JAR3D alignment: Sorted by edit distance

After sorting, we see can easily see that the input sequence is an exact match to loop IL_2AW7_030 in PDB structure 2AW7. An exact sequence match to a loop instance seen in a 3D structure is the best evidence that a sequence can form the 3D geometry seen in a motif group. We can learn more about motif group IL_77895.1 by clicking on the IL_77895.1 link, either on this page or on the JAR3D results page. We can also return to the JAR3D results page by clicking on the query ID link at the top of this page.

JAR3D Query 4295f189-9518-47f9-a4bb-e3d013c3c4f9   +Share
Alignment of Loop 0 to IL_77895.1
Run on Motif Atlas Version 1.18

Input sequences are initially listed on the top, and sequences from 3D structures on the bottom. Click on a column header to sort by that column. Holding shift and clicking on multiple column headers will sort by each column in order.
The strands of your internal loop have been switched to match the strand ordering seen in IL_77895.1. They are still in 5' to 3' order.

Column / Position / Insertion	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Meets Cutoff	Cutoff Score	Interior Edit Distance	Full Edit Distance	Alignment Distance to Sequence0
(Position)	1	1		2		3		4		5				6		7		8		9		10						
(Insertion)	I		I		I		I		I			I		I		I		I		I		I		I				
Sequence0		U		C		A		G		G			*	C		G		C		A		G		True	80.446070	0	0	0
IL_2AW7_030		U		C		A		G		G				C		G		C		A		G						0
IL_4IOA_051		U		G		A		G		G				C		G		C		A		G						1
IL_1S72_023		C		G		A		G		G				C		G		C		A		G						2
IL_3J7A_029		U		C		A		A		G				C		G		U		A		G						2
IL_3U5F_035		U		C		A		A		G				C		G		U		A		G						2
IL_4BPP_035		U		C		A		A		G				C		G		U		A		G						2
IL_4W23_033		U		C		A		A		G				C		G		U		A		G						2
IL_1FJG_023		U		G		A		G		G				U		G		U		A		G						3
IL_2QBG_052		C		G		A		G		G				C		G		U		A		G						3
IL_2QBG_058		U		G		A		G		G				U		G		U		A		G						3
IL_4QCN_049		U		G		A		A		G				C		G		U		A		G						3
IL_3U5F_053		C		G		A		A		G				C		G		U		A		G						4

Persistent link to this alignment page

Motif atlas: Motif group IL_77895.1

As we saw above, motif group IL_77895.1 was a good match to sequence CGCAG*UCAGG. Here is the entry for this motif group in the RNA 3D Motif Atlas. Each motif group consists of instances from high resolution 3D structures of RNA molecules that each have the same basic geometry and the same pattern of basepairing. The conserved pattern of basepairing is indicated by the basepairing diagram in the panel on the left and also in the basepair signature, cWW-tSH-tHW-tHS-cWW, indicating the presence of two flanking cWW basepairs and three non-Watson-Crick basepairs from two different families. See the [RNA Basepair Catalog](#) for basepair examples.
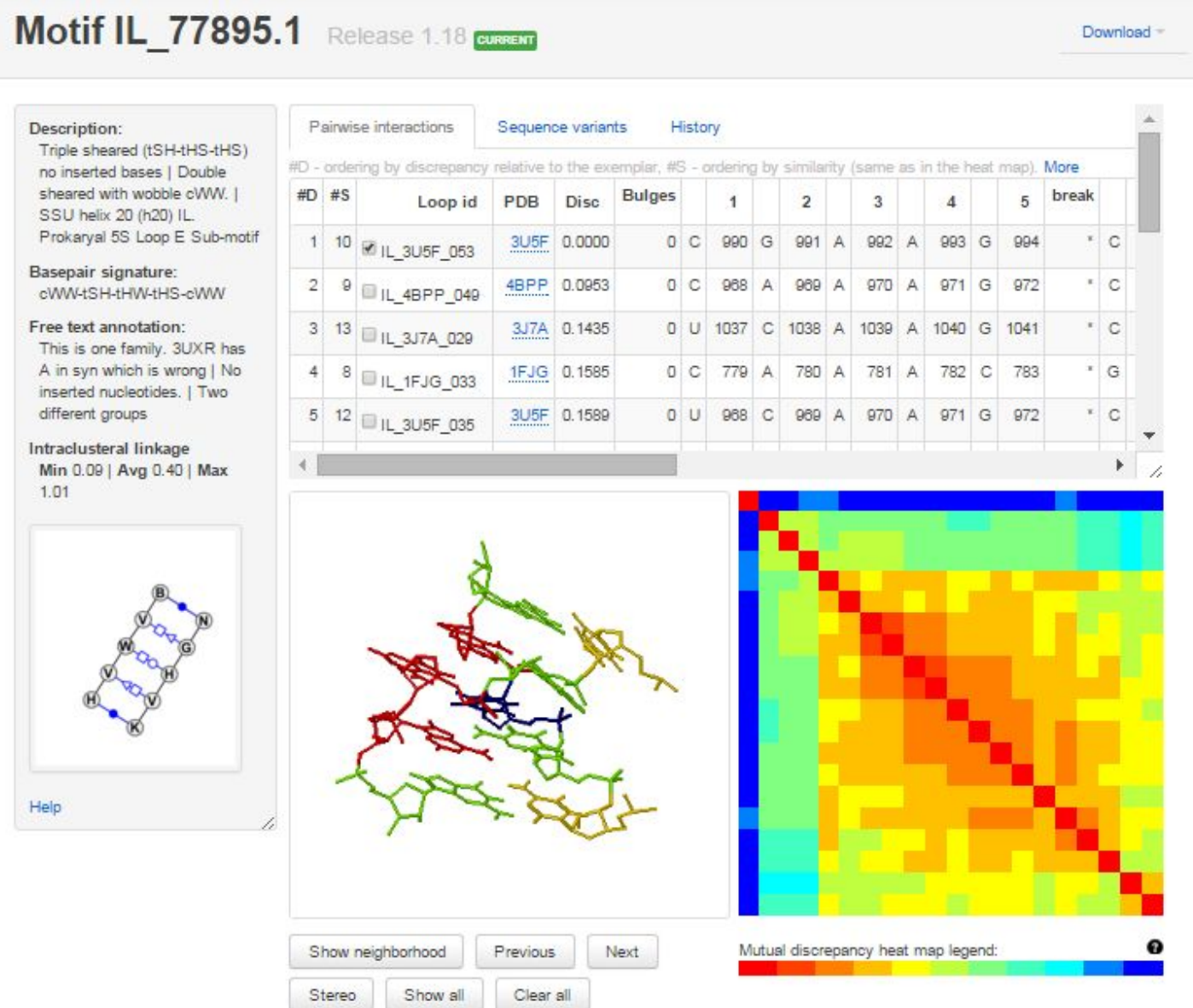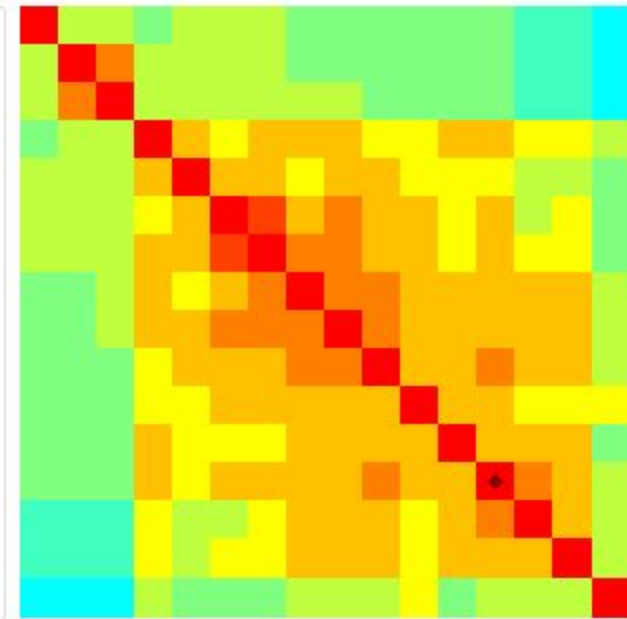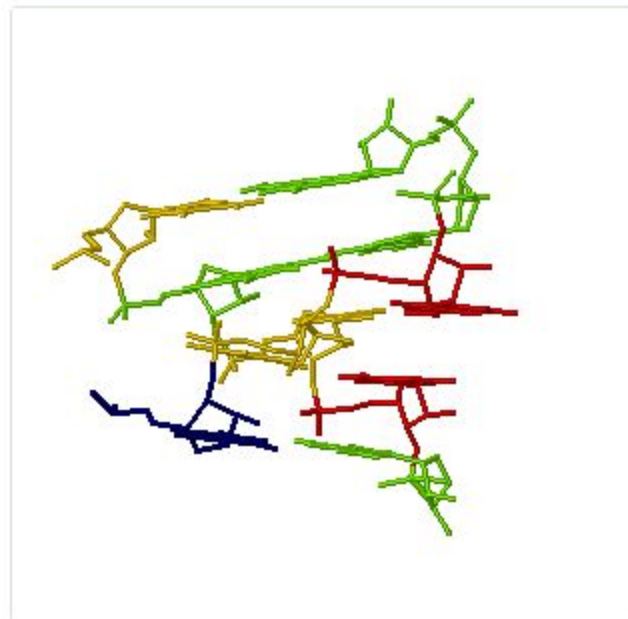

[Link to this page](#)

Motif atlas: Table of instances

The top right of each entry in the Motif Atlas lists instances of that motif taken from RNA 3D structure files of atomic resolution. Each instance occupies one row of the table. The 4-character PDB code for the 3D structure is given. Hovering over the PDB code brings up a short description of the PDB file. Clicking More takes you to the PDB entry for that structure, where you can read all details of the organism, the molecule, the experimental method, etc. The next columns of the table indicate the nucleotide numbers which constitute that instance.

Note that the first instance listed is in fact the internal loop that we read off the secondary structure of the *E. coli* small ribosomal unit. It is also the 3D loop that gave an exact sequence match in the JAR3D alignment page.

The table of instances can be resized by dragging the lower-right corner.

#D	#S	Loop id	PDB	Disc	1	2	3	4	5	6	7	
1	13	☑ IL_2AW7_030	2AW7	0.0000	580 C	581 G	582 C	583 A	584 G	757 U	758 C	759
2	8	☐ IL_3U5F_035	3U5F	0.2193	627 C	628 G	629 U	630 A	631 G	968 U	969 C	970
3	12	☐ IL_2ZJR_050	2ZJR	0.2041	1383 C	1384 G	1385 C	1386 A	1387 G	1370 U	1371 G	1372
4	2	☐ IL_3U5F_053	3U5F	0.5561	1010 C	1011 G	1012 U	1013 A	1014 G	990 C	991 G	992
5	1	☐ IL_1FJG_023	1FJG	0.5531	580 U	581 G	582 U	583 A	584 G	757 U	758 G	759
6	14	☐	2X7M	0.1889	621 C	622 G	623 U	624 A	625 G	946 U	947 C	948

Show neighborhood   Show all

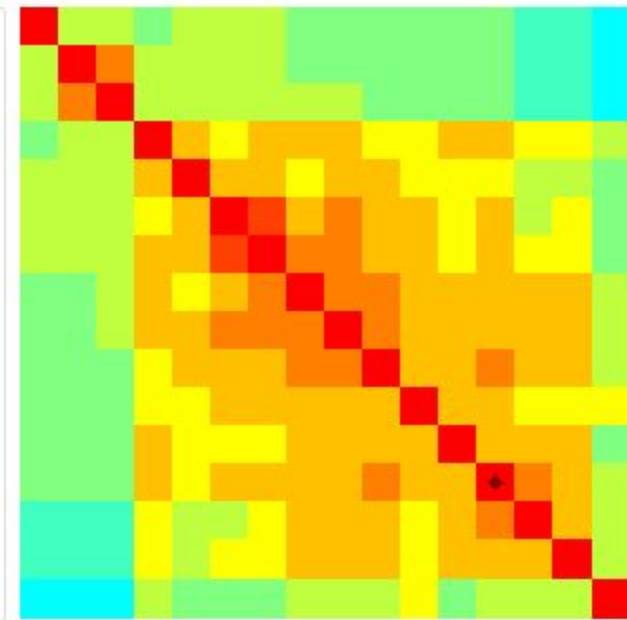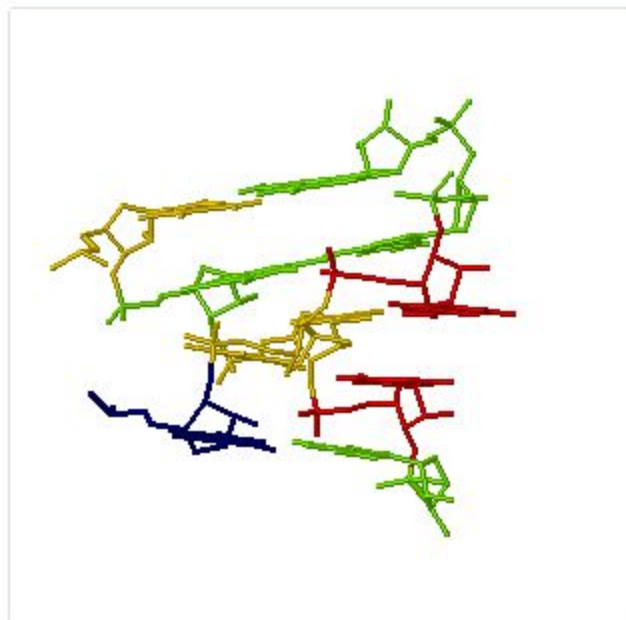Show stereo   Previous   Next   Clear

☐ Nucleotide numbers

Legend: ❓

Link to this page

Motif atlas: Table of instances is an alignment

The table of instances shows an alignment of all 3D instances to one another. The alignment is done in 3D when the RNA 3D Motif Atlas is constructed. This motif group has 10 conserved positions. Some instances may have additional nucleotides which are bulged out and are not core nucleotides of the motif.

Nucleotide numbers increase in each strand of the motif. Position 5 is the last nucleotide of one strand and Position 6 is the first nucleotide of the other strand. Note that in some instances the second strand has higher nucleotide numbers, but in others the second strand has lower nucleotide numbers. One can think of this recurrent RNA internal loop as being a modular construction piece that can be inserted between two helices in one of two orientations, which causes this feature of nucleotide numbers.

#D	#S	Loop id	PDB	Disc	1	2	3	4	5	6	7	
1	13	☑ IL_2AW7_030	2AW7	0.0000	580 C	581 G	582 C	583 A	584 G	757 U	758 C	759
2	8	☐ IL_3U5F_035	3U5F	0.2193	627 C	628 G	629 U	630 A	631 G	968 U	969 C	970
3	12	☐ IL_2ZJR_050	2ZJR	0.2041	1383 C	1384 G	1385 C	1386 A	1387 G	1370 U	1371 G	1372
4	2	☐ IL_3U5F_053	3U5F	0.5561	1010 C	1011 G	1012 U	1013 A	1014 G	990 C	991 G	992
5	1	☐ IL_1FJG_023	1FJG	0.5531	580 U	581 G	582 U	583 A	584 G	757 U	758 G	759
6	14	☐	2X7M	0.1889	621 C	622 G	623 U	624 A	625 G	946 U	947 C	948

Show neighborhood   Show all

Show stereo   Previous   Next   Clear

☐ Nucleotide numbers

Legend: ❓

Link to this page

Motif atlas: Table of instances shows basepairs

Scrolling to the right in the table of instances reveals basepairs made between nucleotides, using the 3-letter Leontis-Westhof abbreviations, n for "near" and a few other annotations such as "perp" for bases that meet at a right angle. Column headers indicate the position of interacting nucleotides; for this motif, they are numbered 1 to 10. Thus the nucleotides in positions 1 and 10 interact to form one flanking Watson-Crick basepair and the nucleotides in positions 5 and 6 interact to form the other flanking Watson-Crick basepair.

Note that the cWW, tSH, tWH, tHS, and cWW basepairs are well conserved down the table of instances.

6	7	8	9	10	1-2	1-9	1-10	2-3	2-7	2-9	3-7	3-8	4-6	4-7	5-6
757 U	758 C	759 A	760 G	761 G			cWW			tSH	ntHW	tWH		tHS	cWW
968 U	969 C	970 A	971 A	972 G			cWW			tSH		tWH		tHS	cWW
1370 U	1371 G	1372 A	1373 G	1374 G			cWW			tSH		tWH		tHS	cWW
990 C	991 G	992 A	993 A	994 G			cWW			tSH		tWH		tHS	cWW
757 U	758 G	759 A	760 G	761 G			cWW			tSH		tWH		tHS	cWW
946 U	947 C	948 A	949 A	950 G			cWW			tSH		ntWH		tHS	cWW
719 C	720 G	721 A	722 G	723 G			cWW			tSH		tWH		tHS	cWW

Show neighborhood   Show all   Show stereo

Previous   Next   Clear
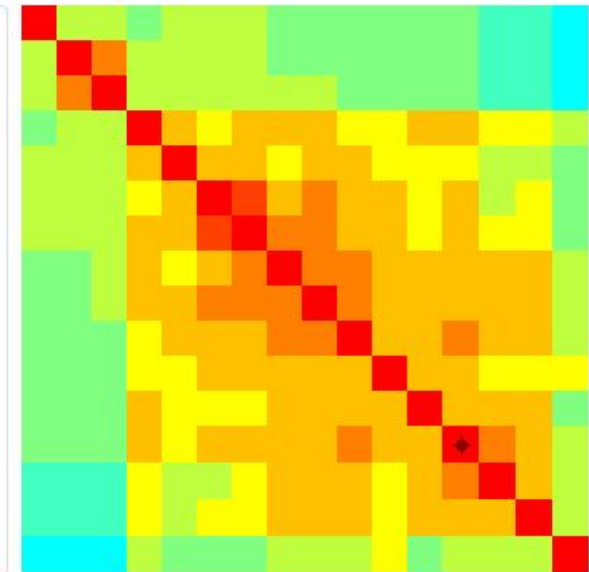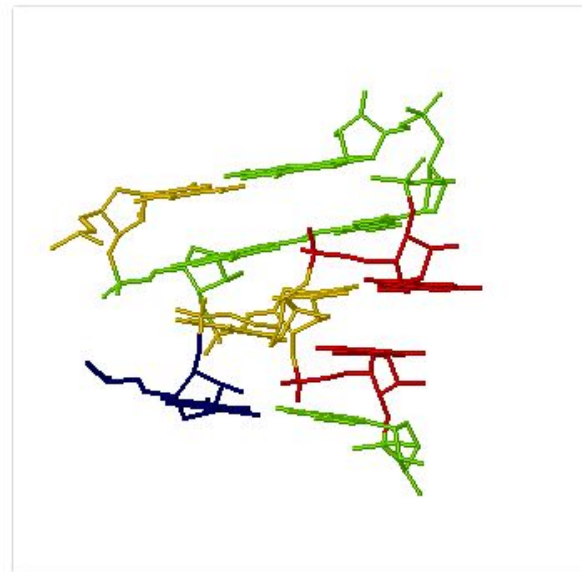
☐ Nucleotide numbers

Legend: ❓
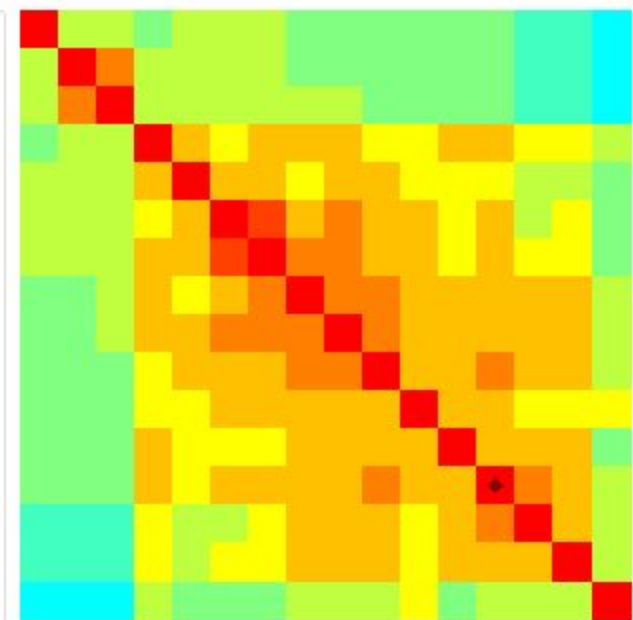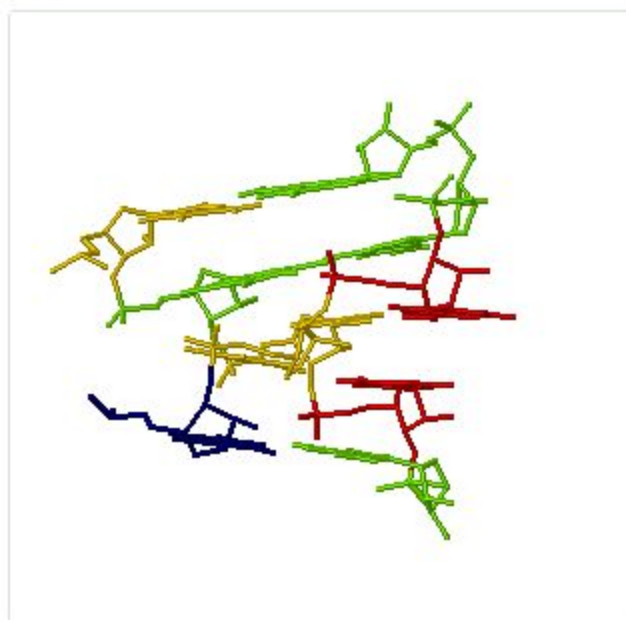
Link to this page

Motif atlas: jsmol window to view instances in 3D

Below the table of instances is a jsmol window which displays the instance or instances selected in the table of instances. Click and drag to rotate the 3D structure. Right click to access additional jsmol capabilities.

Click Next below the jsmol window to see the other instances. Go through the instances and note that they all have the same basic geometry, with some variations. Instances can be superimposed by clicking Show All. Bases are colored red for A, yellow for C, green for G, and blue for U. The superposition shows that most positions have strict conservation, but other positions allow some sequence variability.

#D	#S	Loop id	PDB	Disc	1	2	3	4	5	6	7	
1	13	☑ IL_2AW7_030	2AW7	0.0000	580 C	581 G	582 C	583 A	584 G	757 U	758 C	759
2	8	☐ IL_3U5F_035	3U5F	0.2193	627 C	628 G	629 U	630 A	631 G	968 U	969 C	970
3	12	☐ IL_2ZJR_050	2ZJR	0.2041	1383 C	1384 G	1385 C	1386 A	1387 G	1370 U	1371 G	1372
4	2	☐ IL_3U5F_053	3U5F	0.5561	1010 C	1011 G	1012 U	1013 A	1014 G	990 C	991 G	992
5	1	☐ IL_1FJG_023	1FJG	0.5531	580 U	581 G	582 U	583 A	584 G	757 U	758 G	759
6	14	☐	2X7M	0.1889	621 C	622 G	623 U	624 A	625 G	946 U	947 C	948

Show neighborhood   Show all

Show stereo   Previous   Next   Clear

☐ Nucleotide numbers

Legend:

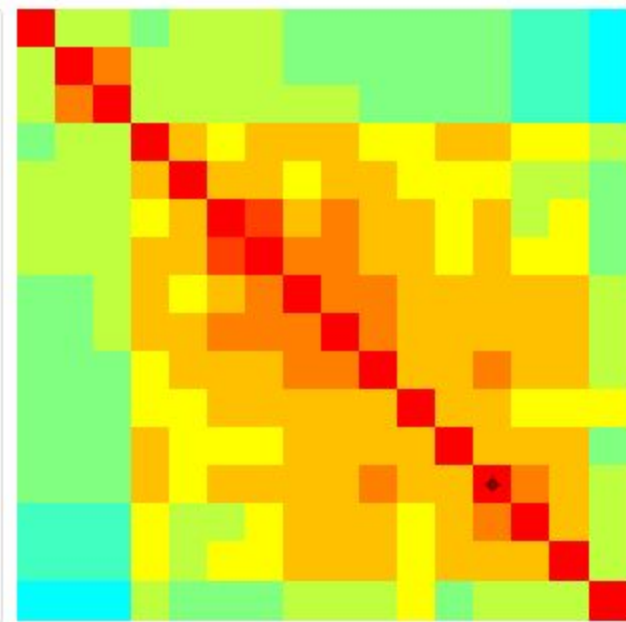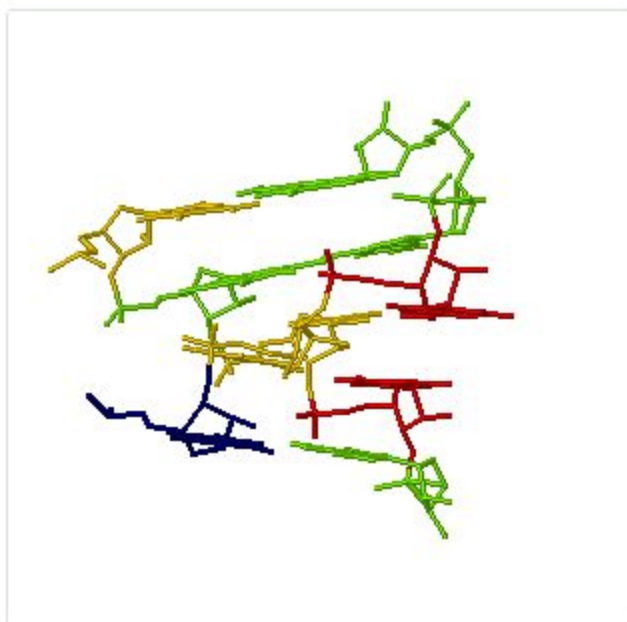# Motif atlas: Mutual discrepancy matrix

To the right of the jsmol window is a square matrix that indicates the geometric discrepancy between each instance and all other instances. This gives an idea how similar or dissimilar all instances are. In this case, all instances have reasonably similar geometries. One can hover over the diagonal of the matrix to see which instance is represented in each row. The instances are not in the same order as in the table above, but by clicking on the header of the table labeled #S, one can get the instances listed in the same order that they appear in the mutual discrepancy matrix.

Click on the diagonal of the matrix to display an instance. Click again to remove it from the display.

Click above the diagonal of the matrix to superimpose two instances, one defined by the row, the other defined by the column.

Click below the diagonal to superimpose ALL instances between the row and the column you click on.

#D	#S	Loop id	PDB	Disc	1	2	3	4	5	6	7	
1	13	☑ IL_2AW7_030	2AW7	0.0000	580 C	581 G	582 C	583 A	584 G	757 U	758 C	759
2	8	☐ IL_3U5F_035	3U5F	0.2193	627 C	628 G	629 U	630 A	631 G	968 U	969 C	970
3	12	☐ IL_2ZJR_050	2ZJR	0.2041	1383 C	1384 G	1385 C	1386 A	1387 G	1370 U	1371 G	1372
4	2	☐ IL_3U5F_053	3U5F	0.5561	1010 C	1011 G	1012 U	1013 A	1014 G	990 C	991 G	992
5	1	☐ IL_1FJG_023	1FJG	0.5531	580 U	581 G	582 U	583 A	584 G	757 U	758 G	759
6	14	☐	2X7M	0.1889	621 C	622 G	623 U	624 A	625 G	946 U	947 C	948

Show neighborhood   Show all

Show stereo   Previous   Next   Clear
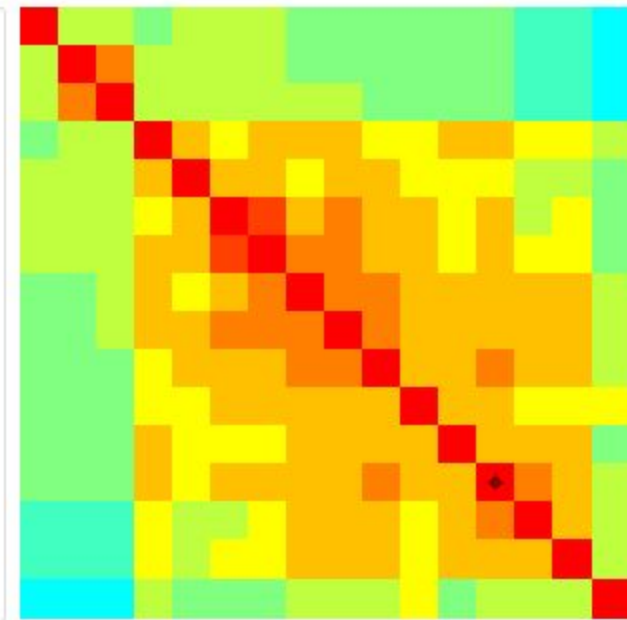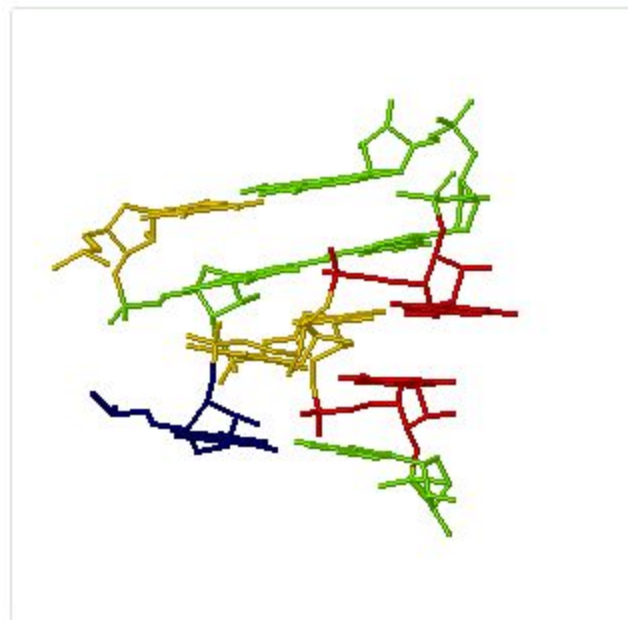
☐ Nucleotide numbers

Legend: ❓

Motif atlas: Interactions with protein chains

Select instance IL_2AW7_030. This instance has the same nucleotide numbers as in the secondary structure earlier in this tutorial, and indeed it is taken from the same molecule, the small ribosomal subunit of E. coli. This is represented by PDB file 2AW7.

Under the jsmol window, click Show Neighborhood to see nearby atoms in the 3D structure file. Nearby RNA nucleotides are shown in a transparent gray, while nearby amino acids are shown in a transparent purple. Hover over a nearby atom to learn its identity. With some effort, one can distinguish the amino acids from the RNA, and so start to see which instances appear to interact with proteins.

By hovering over the amino acids, one can discover exactly which chains and residues make the interaction. Many of the instances in this motif group make substantial interactions with protein chains.

#D	#S	Loop id	PDB	Disc	1	2	3	4	5	6	7	
1	13	☑ IL_2AW7_030	2AW7	0.0000	580 C	581 G	582 C	583 A	584 G	757 U	758 C	759
2	8	☐ IL_3U5F_035	3U5F	0.2193	627 C	628 G	629 U	630 A	631 G	968 U	969 C	970
3	12	☐ IL_2ZJR_050	2ZJR	0.2041	1383 C	1384 G	1385 C	1386 A	1387 G	1370 U	1371 G	1372
4	2	☐ IL_3U5F_053	3U5F	0.5561	1010 C	1011 G	1012 U	1013 A	1014 G	990 C	991 G	992
5	1	☐ IL_1FJG_023	1FJG	0.5531	580 U	581 G	582 U	583 A	584 G	757 U	758 G	759
6	14	☐	2X7M	0.1889	621 C	622 G	623 U	624 A	625 G	946 U	947 C	948

Show neighborhood Show all
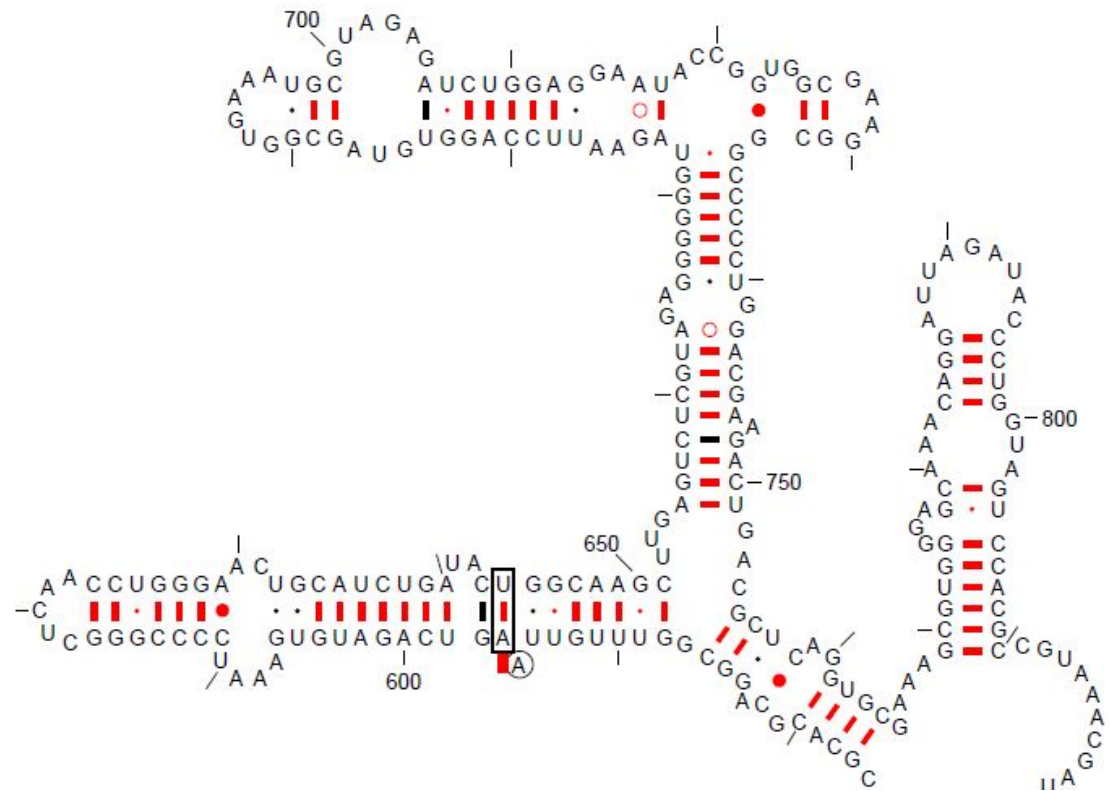
Show stereo Previous Next Clear

☐ Nucleotide numbers

Legend:

[Link to this page](#)

Motif atlas: Two instances of the same motif

Click the PDB heading in the table of instances for motif group IL_77895.1 to sort by PDB file and note that there are two instances from 2AW7, one starting at nucleotide 580, the other at nucleotide 799. The "799" instance is very close to the first instance from E. coli on the secondary structure, and can be found on the secondary structure which you studied before, and which is reproduced here. Find it on the secondary structure, write out its sequence, and then compare with the nucleotide numbers in the table of instances, which are reproduced below.
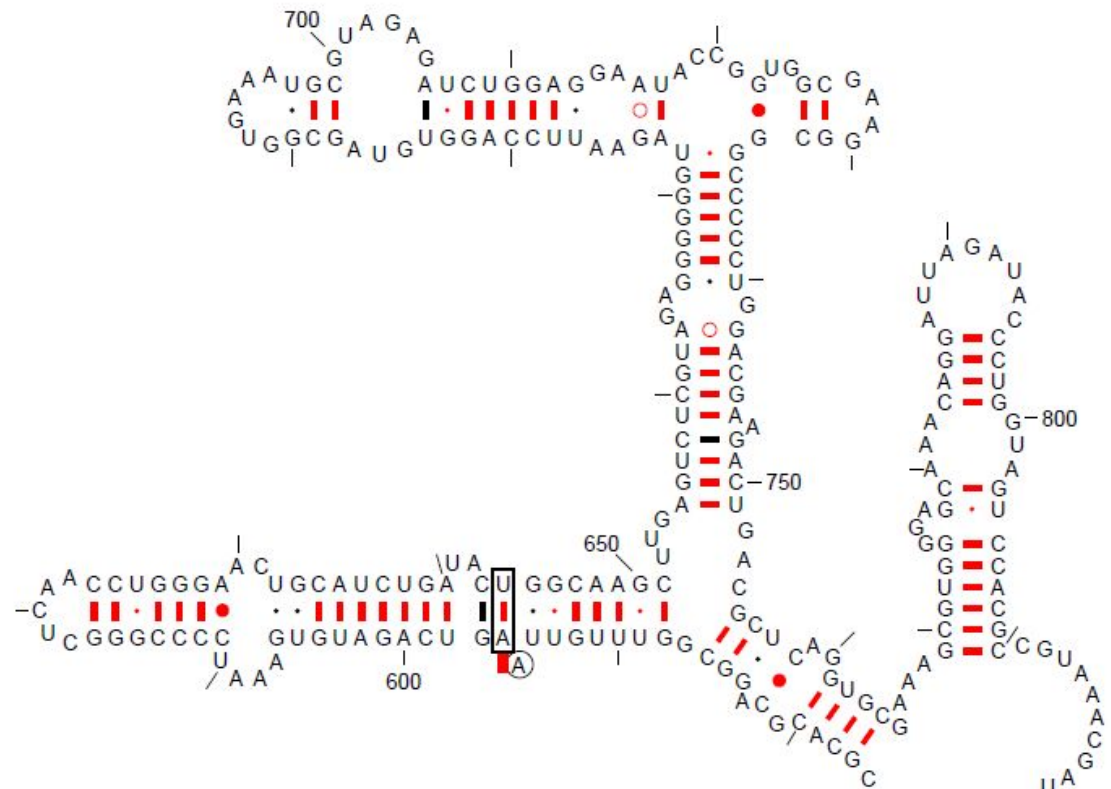

1	13	☑ IL_2AW7_030	2AW7	0.0000	580 C	581 G	582 C	583 A	584 G	757 U	758 C	759 A	760 G	761 G
11	5	☐ IL_2AW7_040	2AW7	0.3043	799 G	800 G	801 U	802 A	803 G	779 C	780 A	781 A	782 A	783 C

Motif atlas: Two instances of the same motif

Something interesting is going on about the order of the nucleotides. Whereas the "580" instance we studied has the nucleotide numbers in the first strand lower than the nucleotide numbers in the second strand, the "799" instance has higher nucleotide numbers on the first strand. In a sense, the internal loop is "rotated" 180 degrees between the two instances. This is a fact of life with internal loops. While the 5′ to 3′ order is very important within each strand, between strands, 5′ and 3′ order is not important. The alignment of instances within this motif group recognizes that.


1	13	☑ IL_2AW7_030	2AW7	0.0000	580 C	581 G	582 C	583 A	584 G	757 U	758 C	759 A	760 G	761 G
11	5	☐ IL_2AW7_040	2AW7	0.3043	799 G	800 G	801 U	802 A	803 G	779 C	780 A	781 A	782 A	783 C

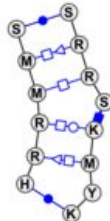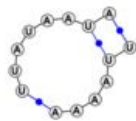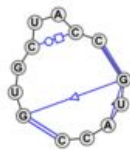# Example: h27 in mitochondrial SSU

Mammalian mitochondrial ribosomes are reduced in size compared to bacterial ribosomes, and they have different base content. Below is a portion of the secondary structure of helix 27 of the Sus scrofa (wild boar) mitochondrial small ribosomal subunit from the Comparative RNA Web (CRW) from this link. This secondary structure was developed before the 3D structure (PDB ID 5AJ3) became available in April 2015. As such, it provides a test of the ability of JAR3D to match sequences to geometries that were known before April 2015.

The secondary structure shows an internal loop with nucleotide ranges 491:498 and 509:514. The numbering scheme is from the Sus scrofa 3D structure 5AJ3. The sequence of the IL is GUACUACU*UAAAAC. We submit this single sequence to JAR3D using release 1.18 of the Motif Atlas, which was generated in December 2014 and so did not include a mitochondrial 3D structure.
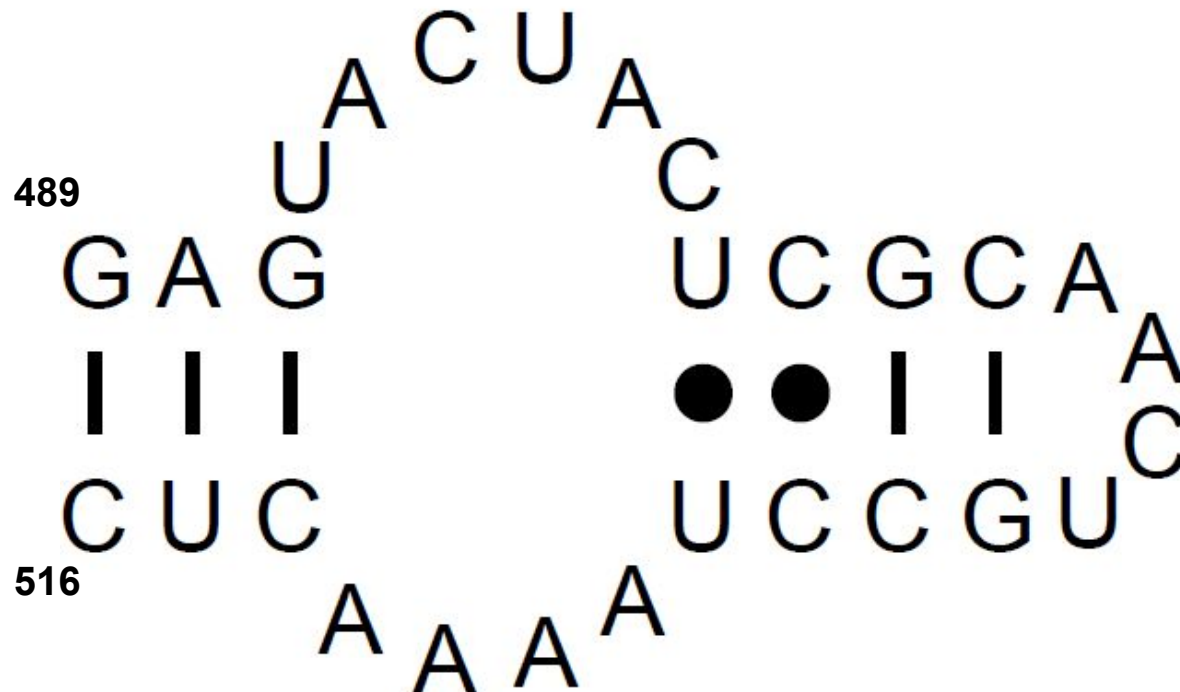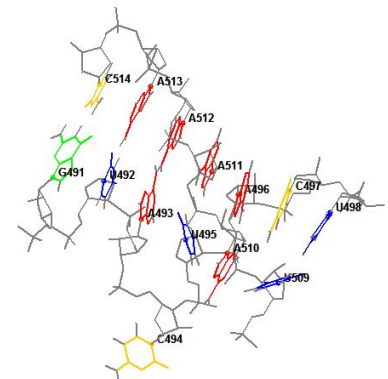
Example: h27 in mitochondrial SSU

Below we show the JAR3D result for the mitochondrial SSU h27 IL sequence GUACUACU*UAAAAC run against release 1.18 of the RNA 3D Motif Atlas, which does not contain a mitochondrial SSU 3D structure. The JAR3D results are available at this persistent link. The first match is solid; the sequence falls in the acceptance region and the cutoff score reasonably far from 0 (positive cutoff scores are in the acceptance region, 100 is the best cutoff score). Motif group IL_95652.6 is a variant of the Sarcin-Ricin internal loop, containing instances from h27 of bacterial and eukaryotic SSU. Note that the sequence from Sus scrofa is distinct from all 3D instances known as of December 2014.

#	Motif ID and Additional Information	Acceptance Rate	Mean Cutoff Score	Full Edit Distance	Interior Edit Distance	2D Diagram
1	◉ IL_95652.6 View IL_95652.6 cWW-tSH-tHH-cSH-tWH-tHS-L-cWW Align sequences to IL_95652.6	100.00	32.19	3	2	
2	○ IL_73656.1 View IL_73656.1 cWW-L-R-R-R-L-L-L-L-cWW-cWW Align sequences to IL_73656.1	100.00	2.55	5	3	
3	○ IL_48045.1 View IL_48045.1 cWW-tSS-L-R-cSS-R-L-tWH-L-L-cWW Align sequences to IL_48045.1	0.00	-1.44	6	4	

We can confirm the match by looking at the 3D structure of the Sus scrofa mitochondrial SSU h27, shown below. It contains the same stacked basepairs as in IL_95652.6 and so has the same basic geometry as the Sarcin-Ricin motif. However, the characteristic GUA base triple in the Sarcin-Ricin is replaced by CUA in which the C is bulged out in Sus scrofa.
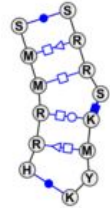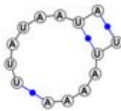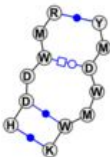
Example: h27 in mitochondrial SSU

Next, we continue the example of h27 in mitochondrial SSU to illustrate the ability of JAR3D to score multiple sequences against known 3D motifs. We use the R3D-2-MSA alignment server to retrieve sequences of mitochondrial SSU h27 from a variety of eukaryotes using the alignment generated by the CRW. Alignment results may be retrieved from this link: http://rna.bgsu.edu/r3d-2-msa?units=5AJ3|1|A||491:5AJ3|1|A||498,5AJ3|1|A||509:5AJ3|1|A||514&aid=2 The link uses nucleotide numbering from the Sus scrofa 3D structure 5AJ3, but the sequences retrieved come from an alignment that predates the 3D structure. Some sequence variants are shown below.

Sequence	Number of Occurences (of 308)	% of Total
GAACUACU,UAAAAC	125	40.5844 %
GUACUACU,UAAAAC	38	12.3377 %
GAACUACG,UAAAAC	31	10.0649 %
GAACUACA,UAAAAC	24	7.7922 %
ACACUACG,UAAAAC	15	4.8701 %
GUACUACC,UAAAAC	10	3.2468 %
ACACUACA,UGAAAC	8	2.5974 %
ACACUACA,UGAAAU	5	1.6234 %
GCACUACU,UAAAAC	4	1.2987 %
GGACUACU,UAAAAC	4	1.2987 %
GUACUACA,CAAAAC	4	1.2987 %

Example: h27 in mitochondrial SSU

Continuing from the previous slide, using the "Download alignment" link in the R3D-2-MSA output and choosing FASTA format returns 308 sequences of h27 from different eukaryotes. Unfortunately, one of them has a non-ACGU character and must be removed from the list. Also, the FASTA format in R3D-2-MSA uses commas to separate the strands whereas JAR3D needs the * character. Making this change and submitting the 307 sequences to JAR3D results in an output page available at this persistent link. Once again, IL_95652.6 is the top match. The acceptance rate over 307 sequences is nearly 100%, the mean cutoff score is well away from 0, and the median minimum edit distances show significant differences from all 3D instances known in December 2014.

#	Motif ID and Additional Information	Acceptance Rate	Mean Cutoff Score	Full Edit Distance	Interior Edit Distance	2D Diagram
1	⦿ IL_95652.6 View IL_95652.6 cWW-tSH-tHH-cSH-tWH-tHS-L-cWW Align sequences to IL_95652.6	99.02	35.28	3	2	
2	◯ IL_73656.1 View IL_73656.1 cWW-L-R-R-L-L-L-L-cWW-cWW Align sequences to IL_73656.1	22.15	-11.84	6	4	
3	◯ IL_06584.1 View IL_06584.1 cWW-bif-tWH-L-R-L-cWW-cWW Align sequences to IL_06584.1	19.87	-10.09	8	4	

Example: h27 in mitochondrial SSU

Continuing from the previous slide, we can align the input sequences to motif group IL_95652.6. This is a very large number of sequences to align and generates a large number of columns. Results are available at this persistent link. By sorting on different columns we can make a few observations: Diagnostic information shows six sequences with a cutoff score less than zero, meaning that they do not match IL_95652.6 well at all, but other sequences with cutoff scores in the range from 20 to 55, which indicate good matches. Interior edit distances range from 1 to 2. Position 10, which is G in bacterial SSU h27, is C in all but four sequences, where it is U. Apparently the GUA base triple that is so common in Sarcin-Ricin motifs is not necessary in the mitochondrial SSU at h27.

JAR3D Query 96fb3f5b-a700-4fde-aa7a-8b33539d7c4e   +Share
Alignment of Loop 1 to IL_95652.6
Run on Motif Atlas Version 1.18

Input sequences are initially listed on the top, with a grey background. Sequences from 3D structures are initially listed on the bottom, with a white background. Click on a column header to sort by that column. Holding shift and clicking on multiple column headers will sort by each column in order.
The strands of your internal loop have been switched to match the strand ordering seen in IL_95652.6. They are still in 5' to 3' order.

Column Position Insertion	1 1 I	2	3 2 I	4 3	5	6 4 I	7	8 5	9	10 6 I	11	12 7 I	13	14 8 I	15	16 9 I	17	18 10 I	19	20 11 I	21	22 12 I	23	24 13 I	25	26 14 I	27	28	29	30	31 I	Meets Cutoff	Cutoff Score	Interior Edit Distance	Full Edit Distance	Alignment Distance to >357.1 Abrothrix jelskii	Alignment Distance to >357.1 Abrothrix jelskii
>357.1 Abrothrix jelskii	U	A	A	A	A	C	*	G	A	A	C	U	A	C	U							True	32.348810	2	3	0	0
>357.1 Abrothrix jelskii	U	A	A	A	A	C	*	G	A	A	C	U	A	C	U							True	32.348810	2	3	0	0
>2159.1 Acerodon celebensis	U	A	A	A	A	C	*	G	U	A	C	U	A	C	U							True	32.194866	2	3	1	1

Example: Archaeal 5S rRNA

To illustrate the ability of JAR3D to split up individual loops from an alignment with a dot-bracket secondary structure, we downloaded the Archaeal 5S rRNA alignment from the CRW website using this link, removed non-ACGU characters and reformatted to put each sequence on a single line, and added a dot-bracket secondary structure following the Haloarcula marismortui (H.m.) 2D structure from this link. We used H.m. because there is a 3D structure of the H.m. LSU with 5S rRNA. Below is an image of the secondary structure. Note that it contains five internal loops and two hairpin loops.
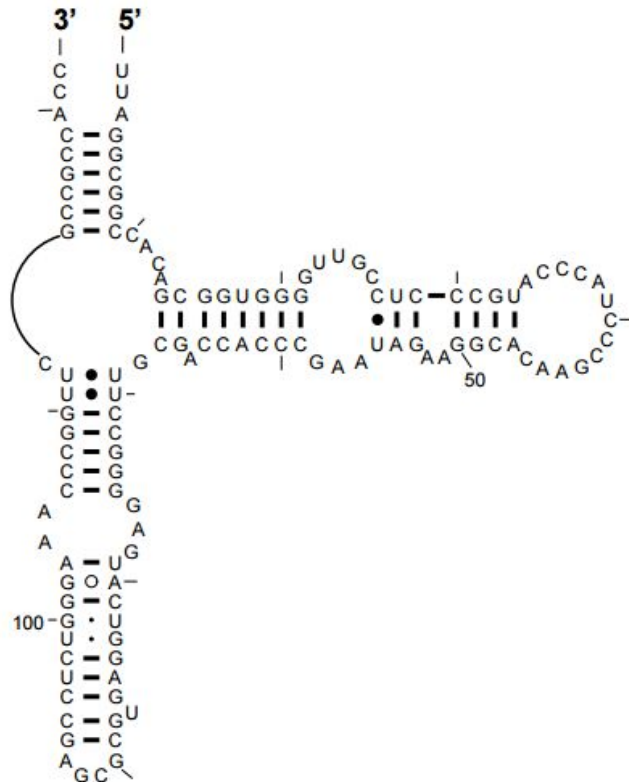
Note that we omitted from the secondary structure the isolated cWW basepair between U79 and A103 shown on the secondary structure below. Isolated cWW basepairs rarely separate real internal loops.

We also omitted from the secondary structure the possible basepair C27 - 55 because it is non-canonical.

Example: Archaeal 5S rRNA

Continuing the previous slide, we submitted the archaeal 5S rRNA with secondary structure to JAR3D. Note that because there are many loops in the structure, it took JAR3D longer than usual to finish processing all loops, so it was necessary to reload the results page once to get results for all of the loops. Results are available at this persistent link, from which the alignment and dot-bracket secondary structure can be obtained.

JAR3D extracted 10 loops, two hairpin loops and eight internal loops. The Input Summary on the JAR3D results page adds column numbers to the input alignment; read these vertically down each column. This makes it easier to identify the different loops. They are labeled on the 2D below.
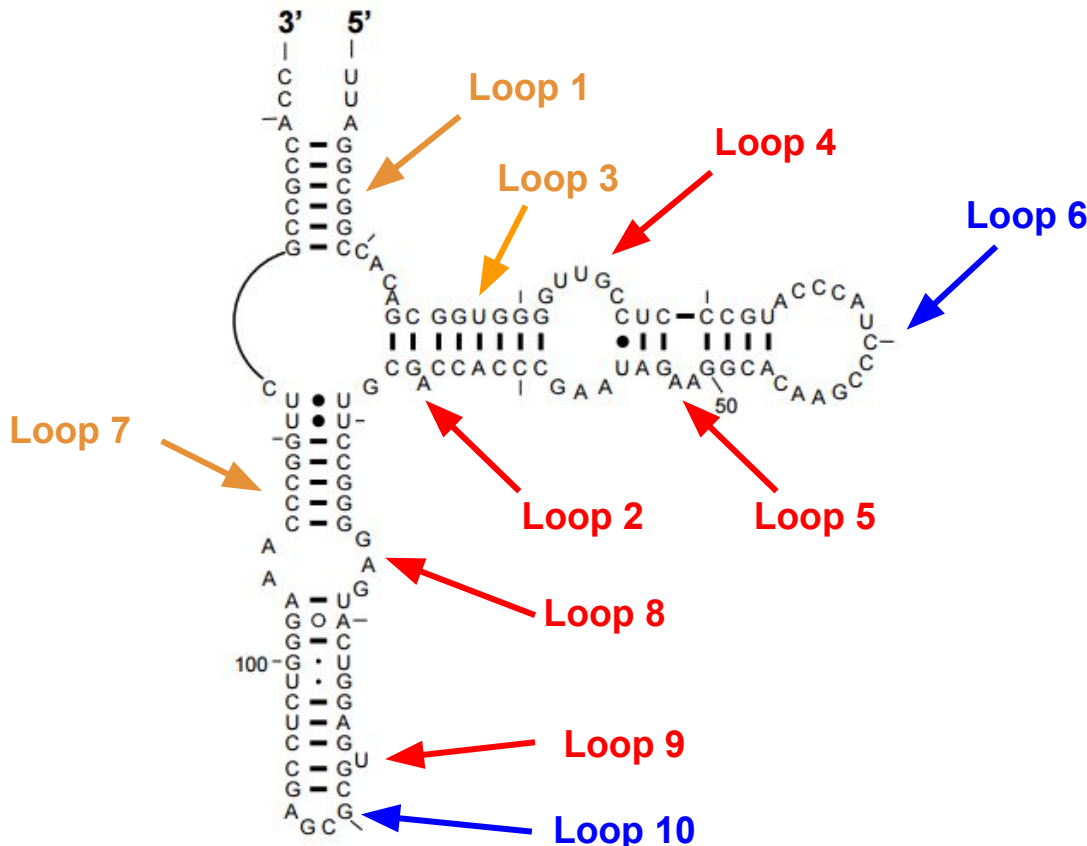

The two hairpin loops are indicated with blue arrows and text. The five internal loops evident in the Haloarcula marismortui secondary structure are indicated with red, and three additional internal loops are indicated in orange.

An inspection of the sequences of Loops 1, 3, and 7 show that they are single base insertions in a very small number of sequences in the alignment, and so will be disregarded from this point forward.

Example: Archaeal 5S rRNA - Loop 4

We continue with the JAR3D results from an archaeal 5S rRNA alignment. Results are available at this persistent link. Loop 4 is a large internal loop. It has only one solid match with JAR3D, to motif group IL_12211.4. This is the motif group that contains the instance of this loop from the H.m. 5S 3D structure, which is found in PDB structure 1S72; the loop ID is 1S72_100. Aligning the sequences of Loop 4 to shows that indeed the sequences from H.m. in the alignment are exact sequence matches to IL_1S72_100. The alignment results are available at this link. A 3D view of the exemplar instance of motif group IL_12211.4 is shown below; it has defined structure but few basepairs.
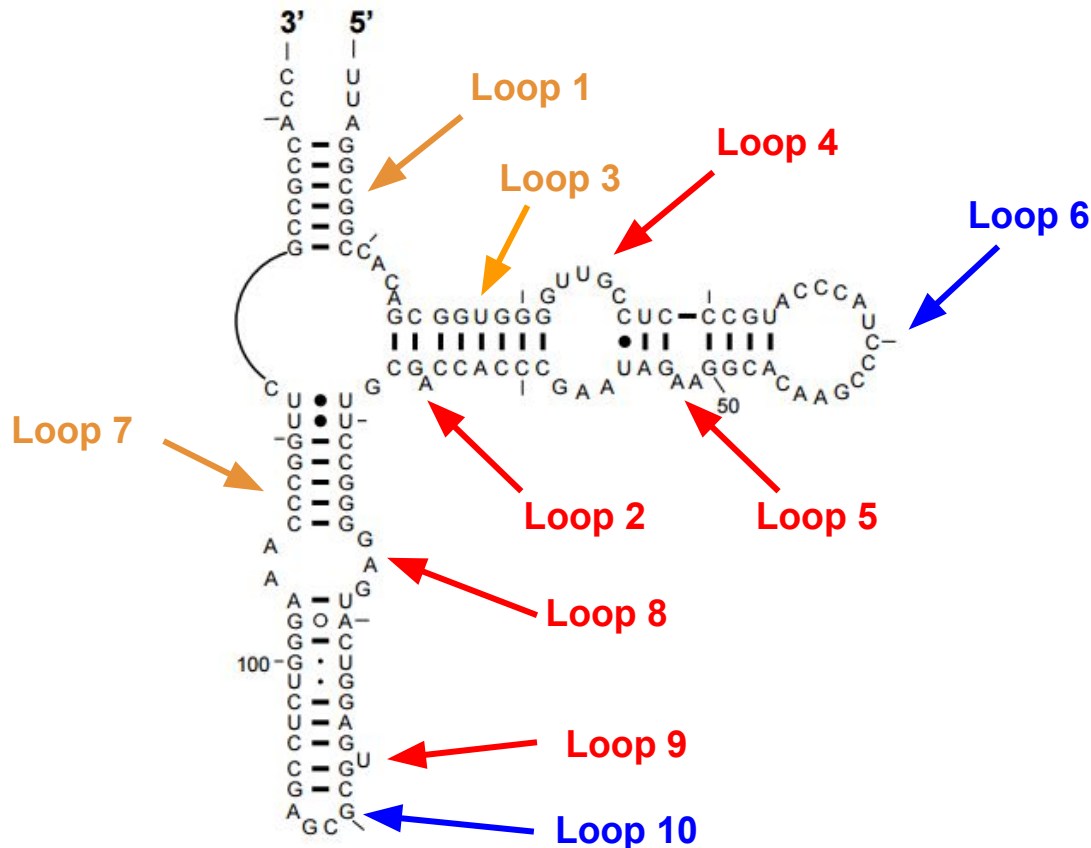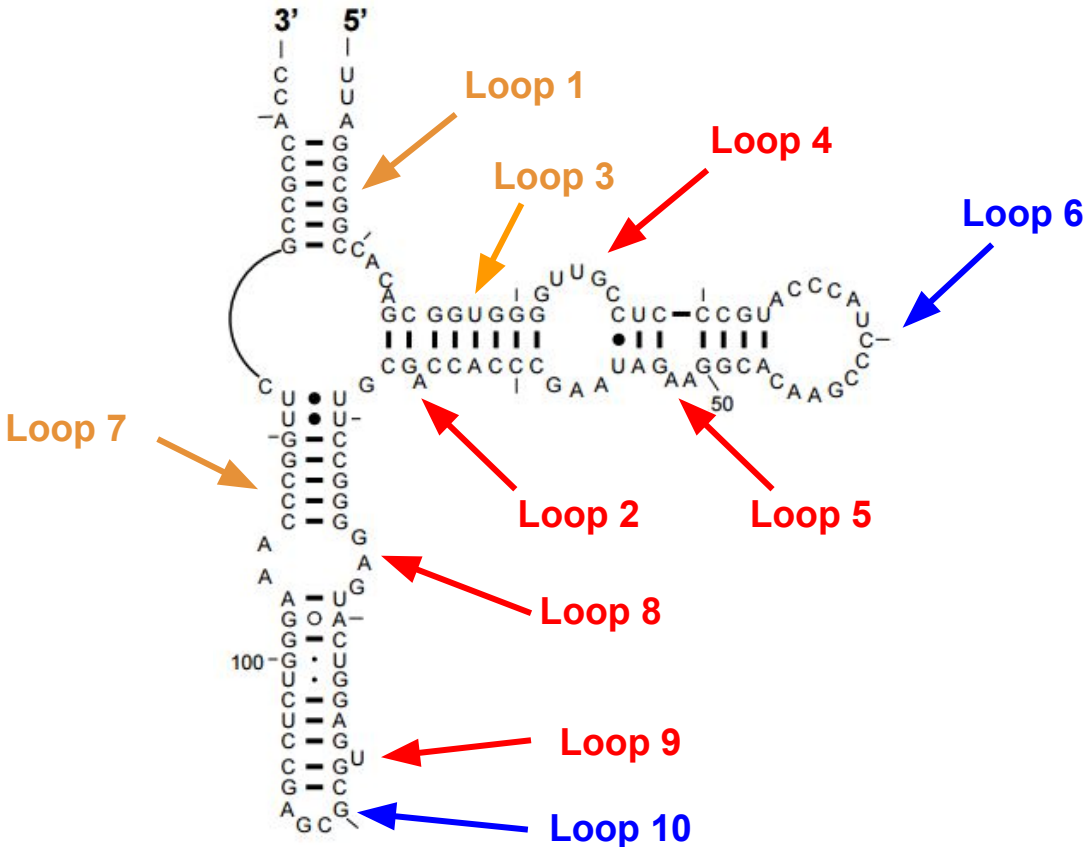

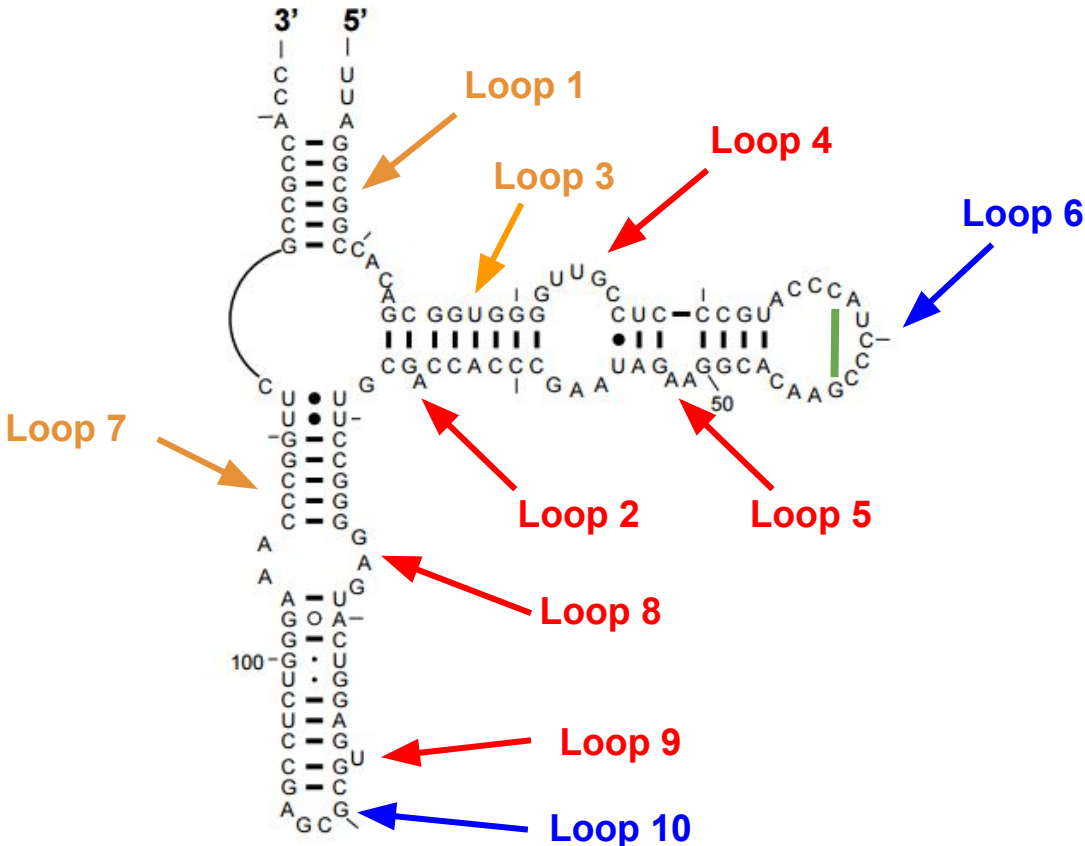Exemplar instance from motif group IL_12211.4:
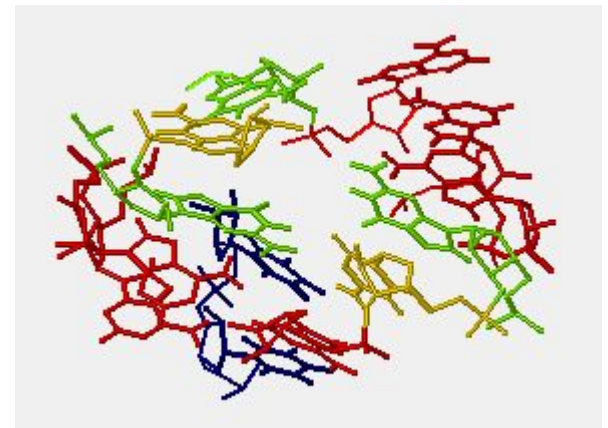
Example: Archaeal 5S rRNA - Loop 5

We continue with the JAR3D results from an archaeal 5S rRNA alignment. Results are available at [this persistent link](#). Loop 5 is a small internal loop which is highly conserved across the archaeal alignment. As such, it is able to match several JAR3D motif groups. In fact, all 10 motif groups listed are reasonable matches to this set of sequences. One lesson from this is that a small internal loop like this has a number of possible geometries, and the actual geometry that it adopts is likely to depend strongly on the local context and additional interactions that it can make in the context of the molecule.
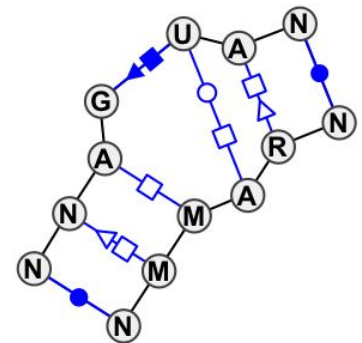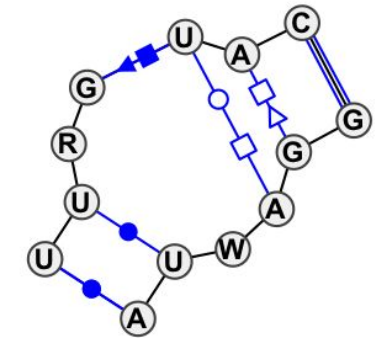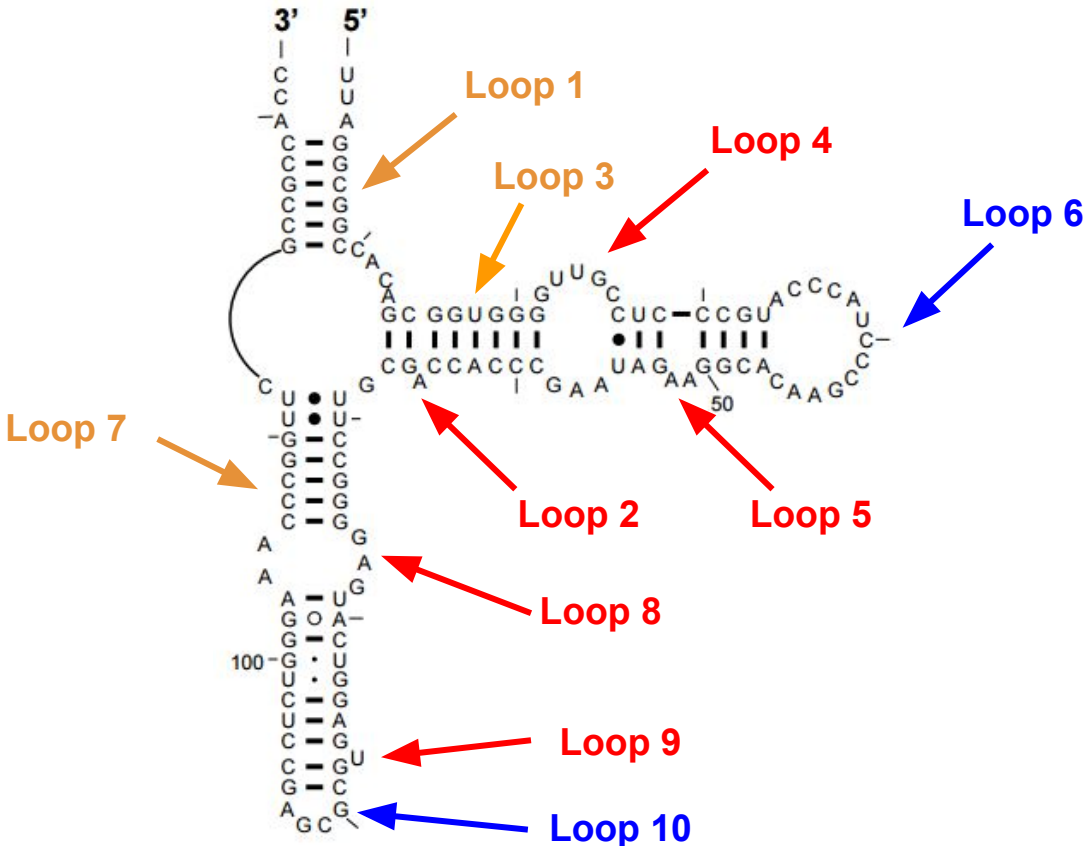
Example: Archaeal 5S rRNA - Loop 6

We continue with the JAR3D results from an archaeal 5S rRNA alignment. Results are available at this persistent link. Loop 6 is a hairpin loop which consistently has 15 nucleotides. All ten motifs listed in the JAR3D results are plausible, but none of them is a good match to the geometry that appears in the known 3D structure of archaeal 5S. This is because of an isolated cWW basepair that occurs between C37 and G43 in the 3D structure, as indicated in green in the secondary structure below. The CG cWW pair shows no sequence variability or covariation over this alignment, so it is no surprise that it was not identified as a cWW basepair. This illustrates a real difficulty in working with RNA internal and hairpin loops; it is necessary to correctly identify the secondary structure first, and that can be very difficult.
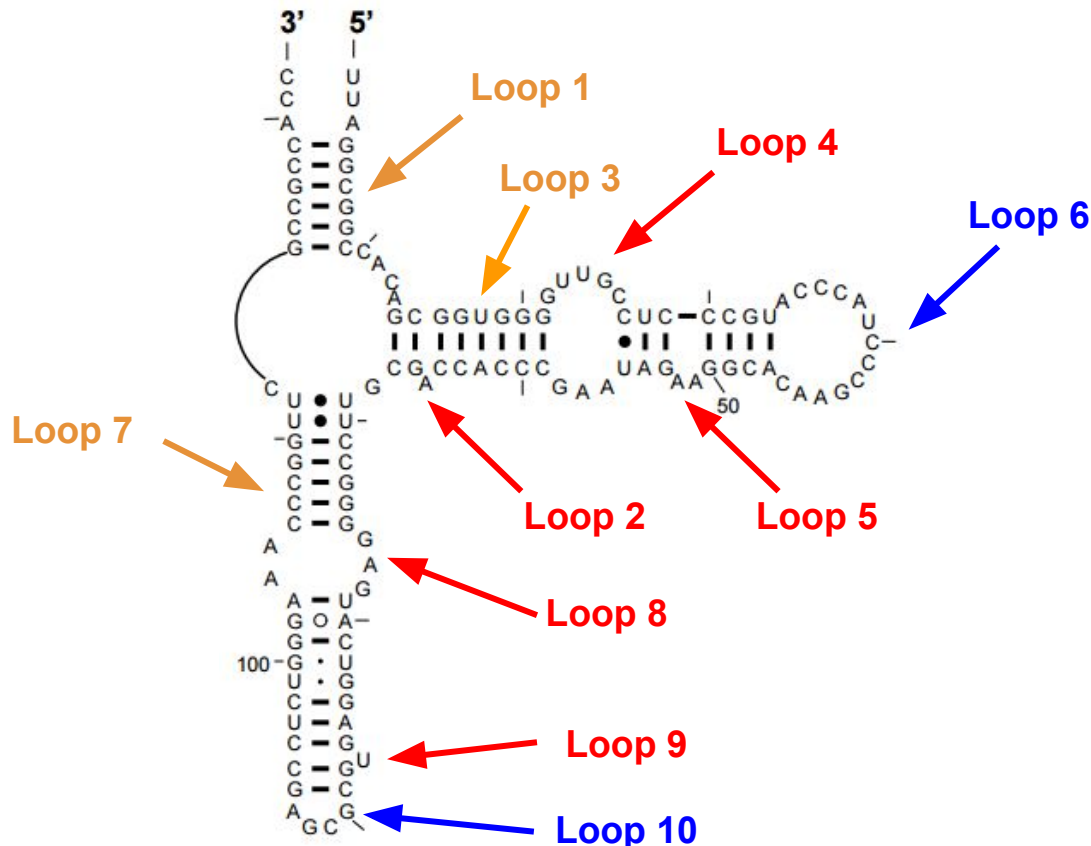
Example: Archaeal 5S rRNA - Loop 8

We continue with the JAR3D results from an archaeal 5S rRNA alignment. Results are available at this persistent link. Loop 8 is a large internal loop. As mentioned above, while the secondary structure below indicates a cWW basepair between U79 and A103, we have omitted it from the dot-bracket secondary structure that we submitted to JAR3D, partly because isolated cWW basepairs rarely occur, but partly because we know ahead of time that the UA do not form a cWW basepair in 3D. JAR3D identifies two good matches, to motif groups IL_21062.1 and IL_49493.8, both of which are Sarcin-Ricin motif groups. The basepair diagrams of these groups are shown below right. They differ only in the basepairs that appear below the GUA base triple. Continued on the next slide.
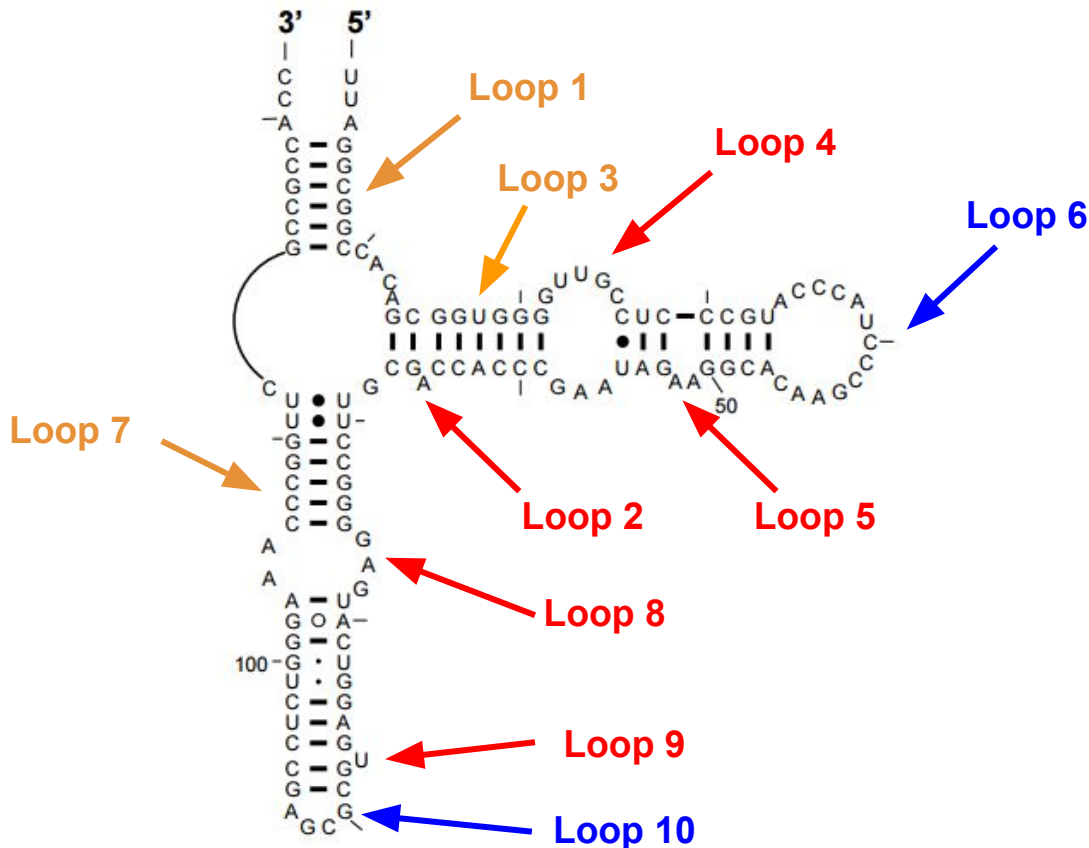
Example: Archaeal 5S rRNA - Loop 8

We continue the discussion of Loop 8. As mentioned on the previous slide, U79-A103 was annotated as making a cWW basepair in the secondary structure. By aligning Loop 8 either to IL_21062.1 or to IL_49493.8, we can see that in the sequence alignment, these bases are very often UA, but occasionally CG, which can explain the inference that it makes a cWW basepair. Most of the sequences are excellent matches to the two motif groups. However, a number of *Sulfolobus* sequences are very poor matches to either motif group, and in fact show a number of stacked CG and GC basepairs and just one non-canonical basepair. This helps to explain the apparent covariation of the U79-A103 basepair, but explains why the covariation is not real.

In most sequences the bases corresponding to U79-A103 apparently make a tWH basepair, but in the Sulfolobus sequences, there appears to be a motif swap where the Sarcin-Ricin motif is completely replaced by another motif that is mostly CG and GC cWW basepairs. This sort of investigation is greatly facilitated by the ability to align the sequences to a dominant motif group; this makes the different sequences stand out clearly.

Example: Archaeal 5S rRNA - Loop 9

We continue with the JAR3D results from an archaeal 5S rRNA alignment. Results are available at this persistent link. Loop 9 is a small internal loop. The bulged U is highly conserved across the archaeal 5S sequence alignment. There are many motif groups which are good matches to this small internal loop, making it very difficult to identify a single geometry consistent with the sequences.

Example: Archaeal 5S rRNA - Loop 10

We continue with the JAR3D results from an archaeal 5S rRNA alignment. Results are available at this persistent link. Loop 10 is a hairpin loop. For reasons that are not clear, the JAR3D results list only one motif group for these sequences. Submitting these sequences to JAR3D separately works just fine, the results just for Loop 10 are available at this persistent link. Once again, all ten motif groups listed are good matches, and most of them are variations on the GNRA hairpin in which the GA make a tSH basepair with the N and R bases stacked on top of this. In fact the H.m. 3D structure in 1S72 has a hairpin that is grouped with motif group HL_67042.17, a large motif group containing many instances. An alignment of the Loop 10 sequences to this motif group is available at this link.
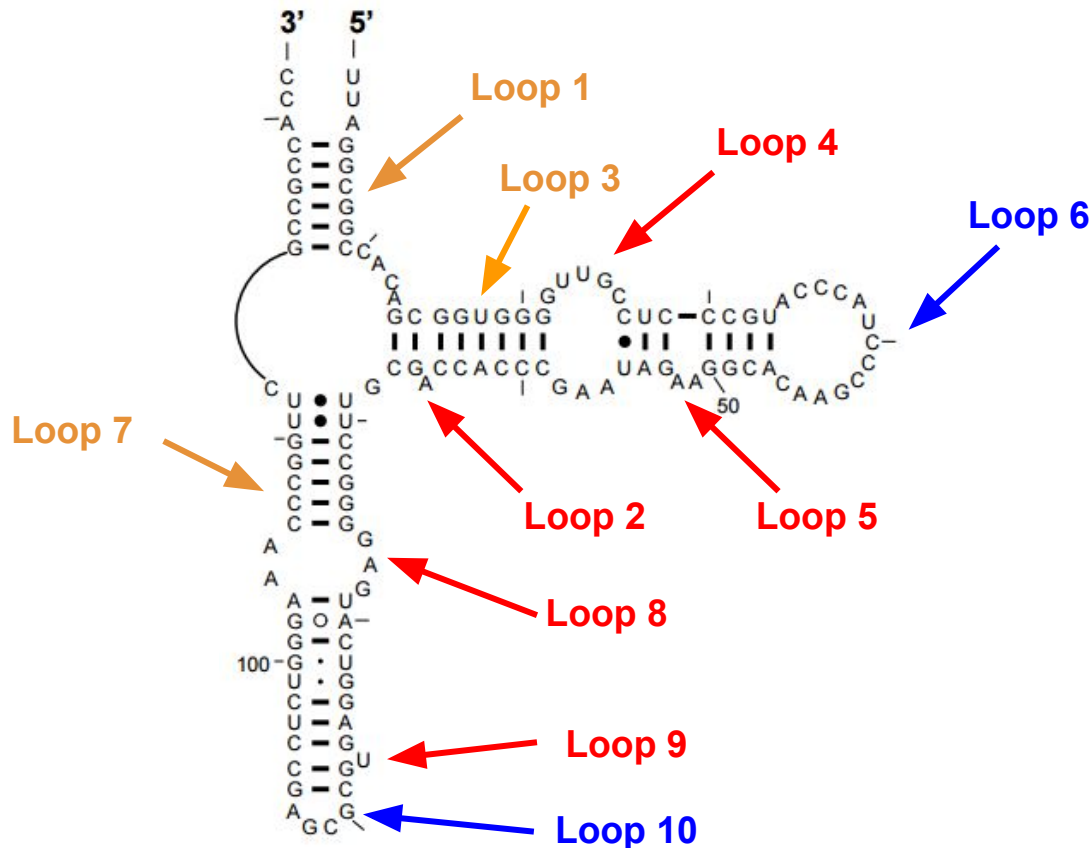
As with Loop 8, a small number of sequences do not match the pattern of the GNRA hairpin well at all. This is fairly common with GNRA hairpins. While the GNRA geometry is known to facilitate long-range RNA-RNA interactions, some organisms lack the GNRA geometry; perhaps their RNA molecules and associated proteins have managed to find a new way to make the long-range interactions.

Overview of resources and tools

- [Non-redundant sets](#) - Groups all RNA 3D structure files into equivalence classes and identifies representative structures.
- [RNA basepair catalog](#) - Shows examples of each base combination in each basepairing family and a measure of isostericity between all base combinations
- [RNA 3D Motif Atlas](#) - Loops collected from all high-resolution RNA 3D structures, organized into coherent groups
- [JAR3D](#) - Submit sequence(s) of an RNA loop, JAR3D will score the sequence(s) against all motif groups in the RNA Motif Atlas
- These and more are available from our group's main website: [rna.bgsu.edu](#)

Acknowledgments

Collaborators:

Craig L. Zirbel, BGSU Department of Mathematics and Statistics

Neocles Leontis, BGSU Department of Chemistry

Eric Westhof, IBMC Strasbourg, Université de Strasbourg, Strasbourg, France

Jiri and Judit Sponer Masaryk, Univ. Brno, Czech Republic

Helen Berman and John Westbrook, Rutgers University and PDB/NDB

Recent and Current Students:

- Dr. Jesse Stombaugh
- Dr. Ali Mokdad
- Dr. Michael Sarver
- Dr. Ryan Rahrig
- Dr. Amal Abu Almakarem
- Dr. Anton Petrov
- Dr. Emil Khisamutdinov
- Jim Roll
- Blake Sweeney
- Maryam Hosseini
- Poorna Roy
- Sri Devan
- Abdul Al Nimer